Supplementary Information for

An additional water is introduced into the manganese cluster during the formation of the S_3 state of Photosystem II

Asmit Bhowmick[1#], Miao Zhang[1#], Philipp S. Simon[1#], Hiroki Makita[1#], Isabela I. Nangca[1], Erzsi Szilagyi[1], Moritz Kretzschmar[1,2], Margaret D. Doyle[1], Natalie M. Minnetian[1], Rana Hussein[2,3], Olli Hart[2], Kuntal Chatterjee[1∅], A. Orkun Aydin[4,5], Dmitry Shevela[6], Mun Hon Cheah[4], Nicholas Croy[4], Petko Chernev[4⊗], Thomas Fransson[7], Vandana Tiwari[8], Humberto Sanchez[8], Pamela Schleissner[8], Randy Lemons[8], Greg Gate[8], Meredith Henstridge[8], James M. Glownia[8], Frédéric Poitevin[8], Daniel J. Rosenberg[8], Sebastian Dehe[8], Leland B. Gee[8], Kensuke Tono[9,10], Shigeki Owada[9,10], Roland Oggenfuss[11], Dmitry Ozerov[11], Mathias Sander[11], Roman Mankowsky[11], Henrik T. Lemke[11], Iris D. Young[1+], James M. Holton[1,12], David W. Mittan-Moreau[1], Daniel W. Paley[1], Pavel V. Afonine[1], Nigel W. Moriarty[1], Paul D. Adams[1,13], Fikret Mamedov[4], Holger Dobbek[2], Athina Zouni[2], Roberto Alonso-Mori[8], Uwe Bergmann[14], Aaron S. Brewster[1], Nicholas K. Sauter[1], Johannes Messinger[4,5*], Jan F. Kern[1*], Vittal K. Yachandra[1*], Junko Yano[1*]

[1]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

[2]Humboldt-Universität zu Berlin, Department of Biology, 10099 Berlin, Germany.

[3]Institute of Biochemistry, Biocenter, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany

[4]Molecular Biomimetics, Department of Chemistry-Ångström, Uppsala University, SE 75120 Uppsala, Sweden.

[5]Department of Plant Physiology, Umeå Plant Science Centre, Umeå University, SE 90187 Umeå, Sweden

[6]Department of Chemistry, Umeå University, SE 90187 Umeå, Sweden.

[7]Department of Theoretical Chemistry and Biology, KTH Royal Institute of Technology, 114 28 Stockholm, Sweden

[8]Linac Coherent Light Source, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA.

[9]Japan Synchrotron Radiation Research Institute, Hyogo, Japan.

[10]RIKEN SPring-8 Center, Hyogo, Japan.

[11]SwissFEL, Paul Scherrer Institut, 5232 Villigen, Switzerland

[12]Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, California, CA 94158, USA.

[13]Department of Bioengineering, University of California, Berkeley, CA 94720, USA.

[14]Department of Physics, University of Wisconsin–Madison, Madison, WI 53706, USA.

[+]Present address: Institute of Nanostructure and Solid State Physics, University of Hamburg, 22761 Hamburg, Germany

[⊗]Present address: Tandem Laboratory, Department of Physics, Uppsala University, SE 75120 Uppsala, Sweden

[∅] Present address: Lam Research Corporation, Tualatin, OR, USA

[#]These authors contributed equally.

[*]Correspondence should be addressed to johannes.messinger@umu.se, jfkern@lbl.gov, vkyachandra@lbl.gov, jyano@lbl.gov

Supplementary Information

Supplementary Figures

Supplementary Tables

Supplementary Notes

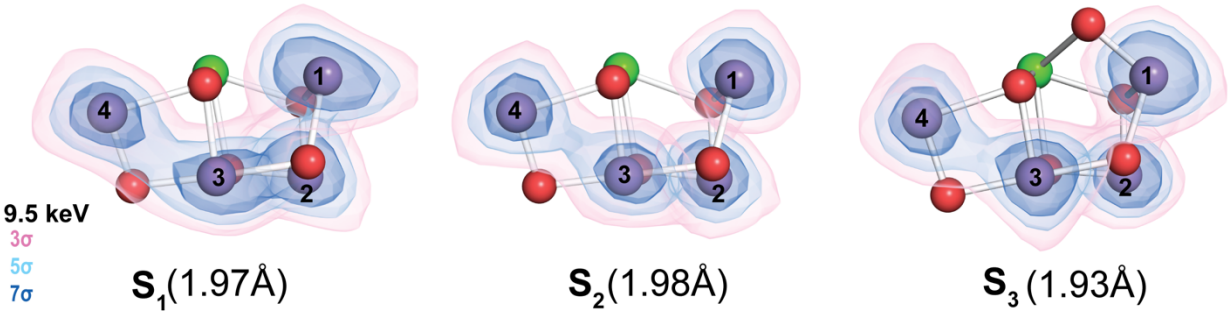
Supplementary Methods

Additional material for paper

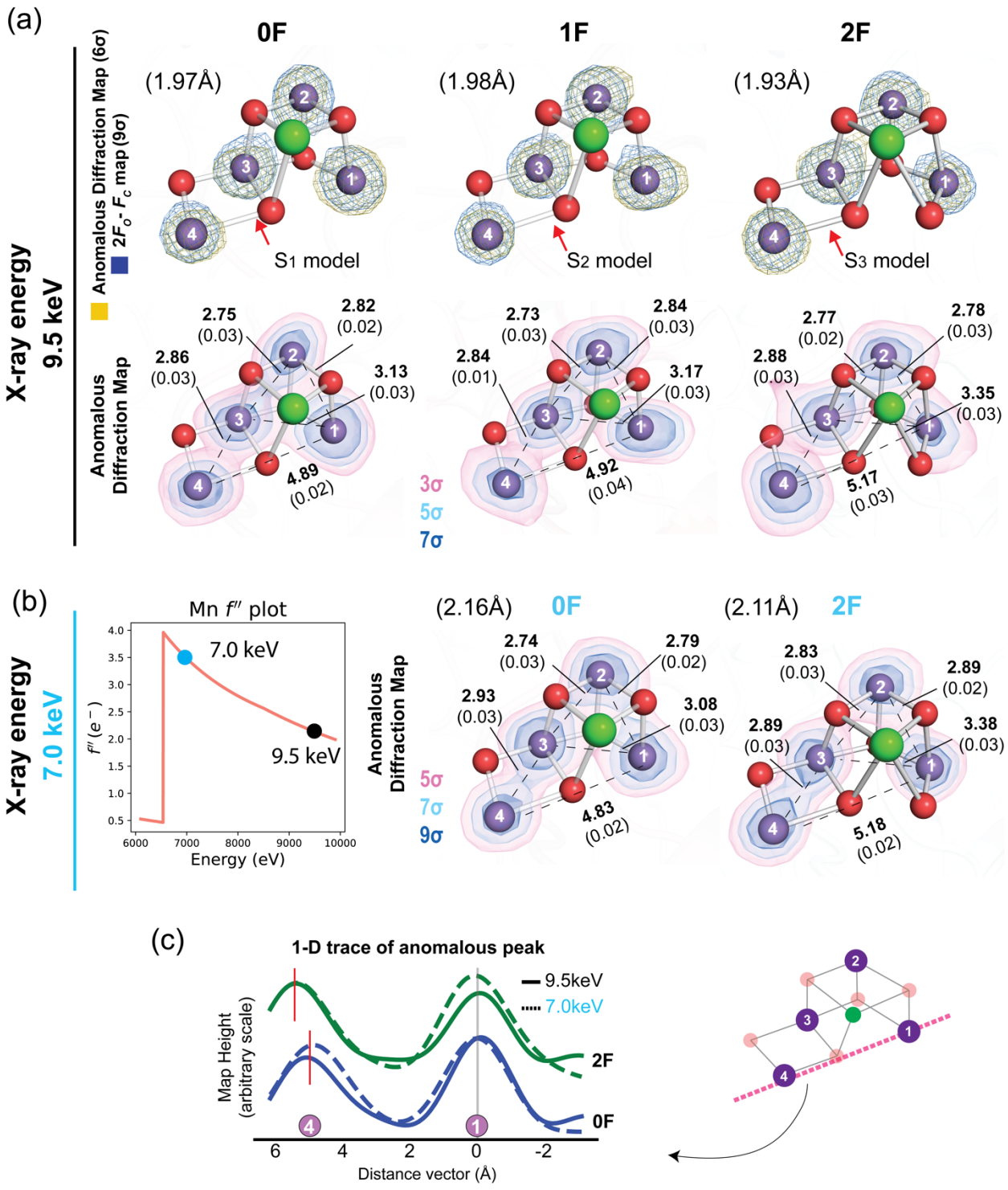
Provided as separate Supplementary Data
File 1: Custom OEC/OEI restraints file (HR restraints)
File 2: Edits file for ligand environment and Mn-Mn distances for the 2F dataset
File 3: phenix.refine settings file for the 2F dataset
File 4: FBHR file for the 2F dataset

Scripts on Github
See Code Availability section -
1. Scripts for calculating anomalous peak 1D traces in Figure 2/ Supplementary Figure 2
2. Scripts for $2F_O$-F_C map 1D traces in Figure 3c/Supplementary Figure 3c
3. Script for computing Restraints report
4. Script for computing anomalous and $2F_O$-F_C centroids (Supplementary Table 5)
5. Script to generate different restraints residuals (Supplementary Figure 13)

Supplementary Figures

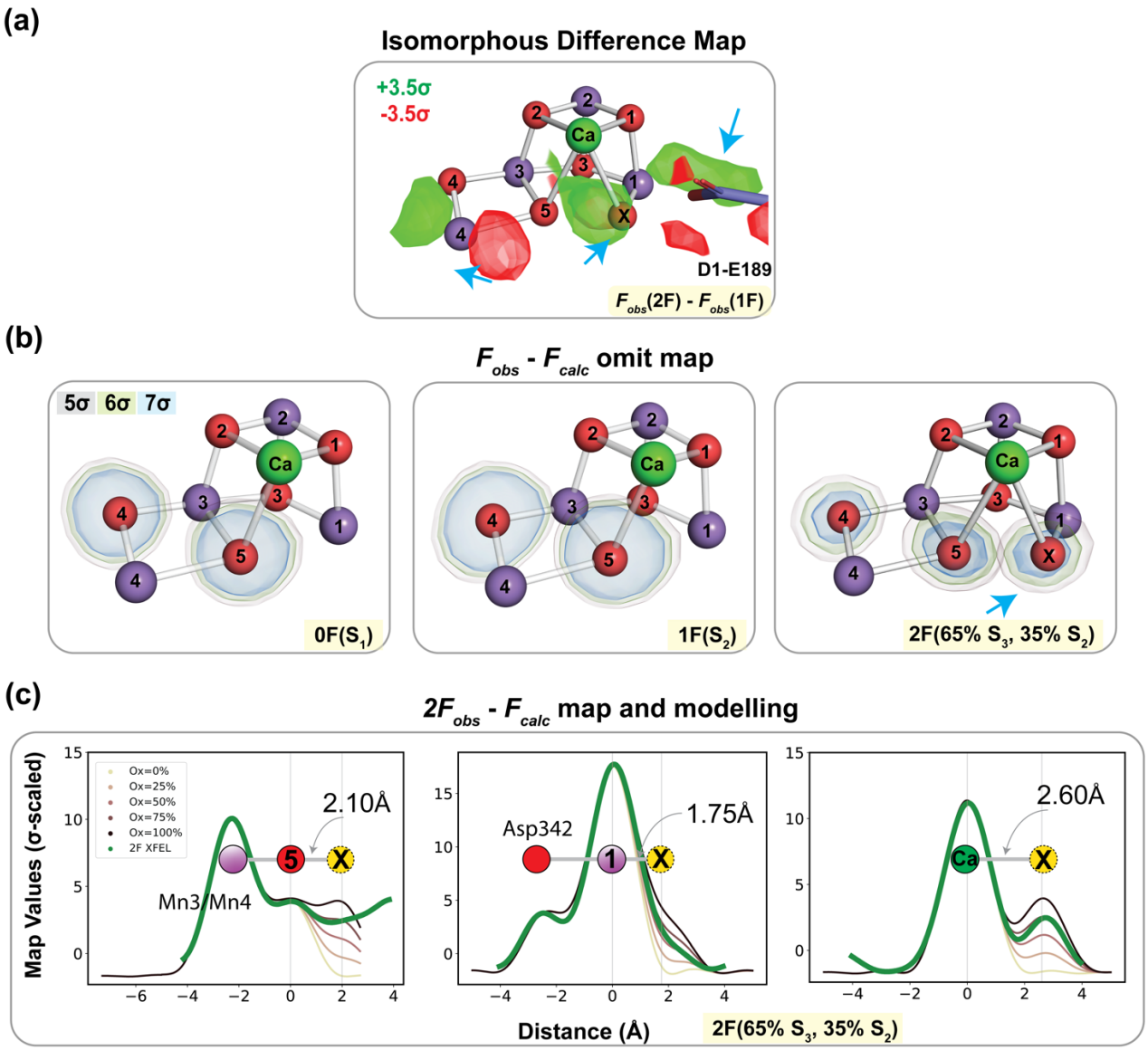


Supplementary Figure 1: (a) Alternate view of the anomalous density around the Mn_4Ca cluster in monomer I for the 9.5 keV data



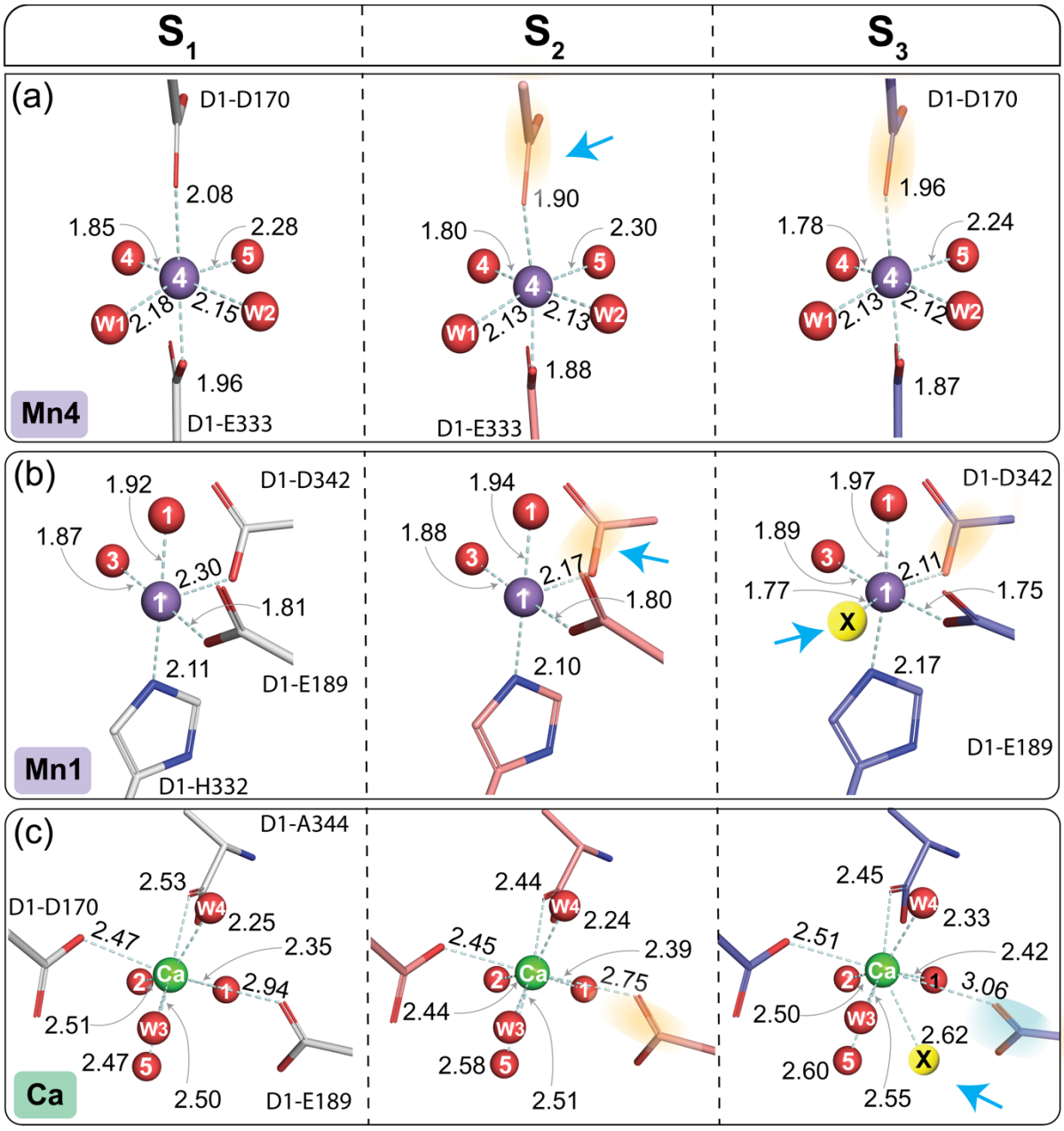
Supplementary Figure 2: Accurate Mn positions in PS II with Anomalous diffraction data and evidence for Mn_4Ca cluster expansion in S_3 state in monomer II (a) *Upper panel*: Overlay of anomalous diffraction maps at 9.5 keV and standard $2F_O$-F_C maps highlighting the degree of similarity of the two maps in all 3 S-states. *Lower panel*: Anomalous diffraction maps at 9.5 keV

for the Mn_4Ca cluster in the 0F (S_1), 1F (S_2) and 2F (S_3 majority) states. Distances shown are for the model refined with anomalous intensities and error estimates are from the END/RAPID procedure. The Mn-Mn distances were subsequently restrained tightly to the values from the anomalous data when we modelled oxygen positions with conventional intensity data. (b) Anomalous diffraction maps at 7.0 keV for the Mn_4Ca cluster in the 0F (S_1) and 2F (S_3 majority) states. Distances shown are for the model refined with anomalous. At 7.0 keV, the f" value of Mn is approximately 3.5 e^- (c) 1-d trace of the anomalous diffraction map along the Mn1-Mn4 vector with the origin being the modelled Mn1 position. A shift in the relative Mn4 position can be seen in the peak confirming expansion of the Mn_4Ca cluster in the S_3 state. Source data for panel c is provided in the Source Data file.

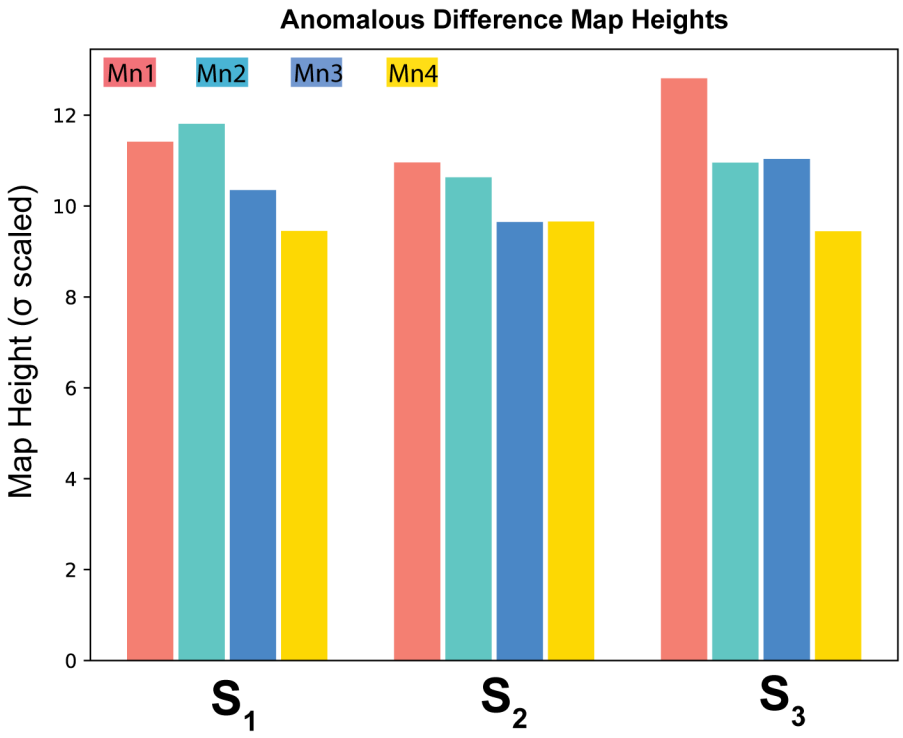


Supplementary Figure 3: Evidence for insertion of O_X in the S_3 state in monomer II: (a) Isomorphous difference maps (F_O - F_O) for the 2F - 1F dataset contoured at +/- 3.5σ. The O_X signal shows up as a 5σ peak in the 2F - 1F difference map. Overlaid is the S_3 structure. (b) F_O - F_C omit maps for the O4, O5 and O_X atoms in the S-states with the O_X omit peak height being 8.5σ. Overlaid are the S_1, S_2 and S_3 structures. (c) 1-d traces of the 2F_O - F_C experimental map for the 2F state along the bond vectors (O5-O_X, Mn1-O_X, Ca-O_X) and comparison with theoretical X-ray scattering map values from 2 or 3-atom models. In the simplified models, the O_X population was varied between 0-100%. An O_X population between 50-75% in the theoretical curves best match the 2F state experimental profiles, supporting its modelling. A more precise estimate of 65% O_X

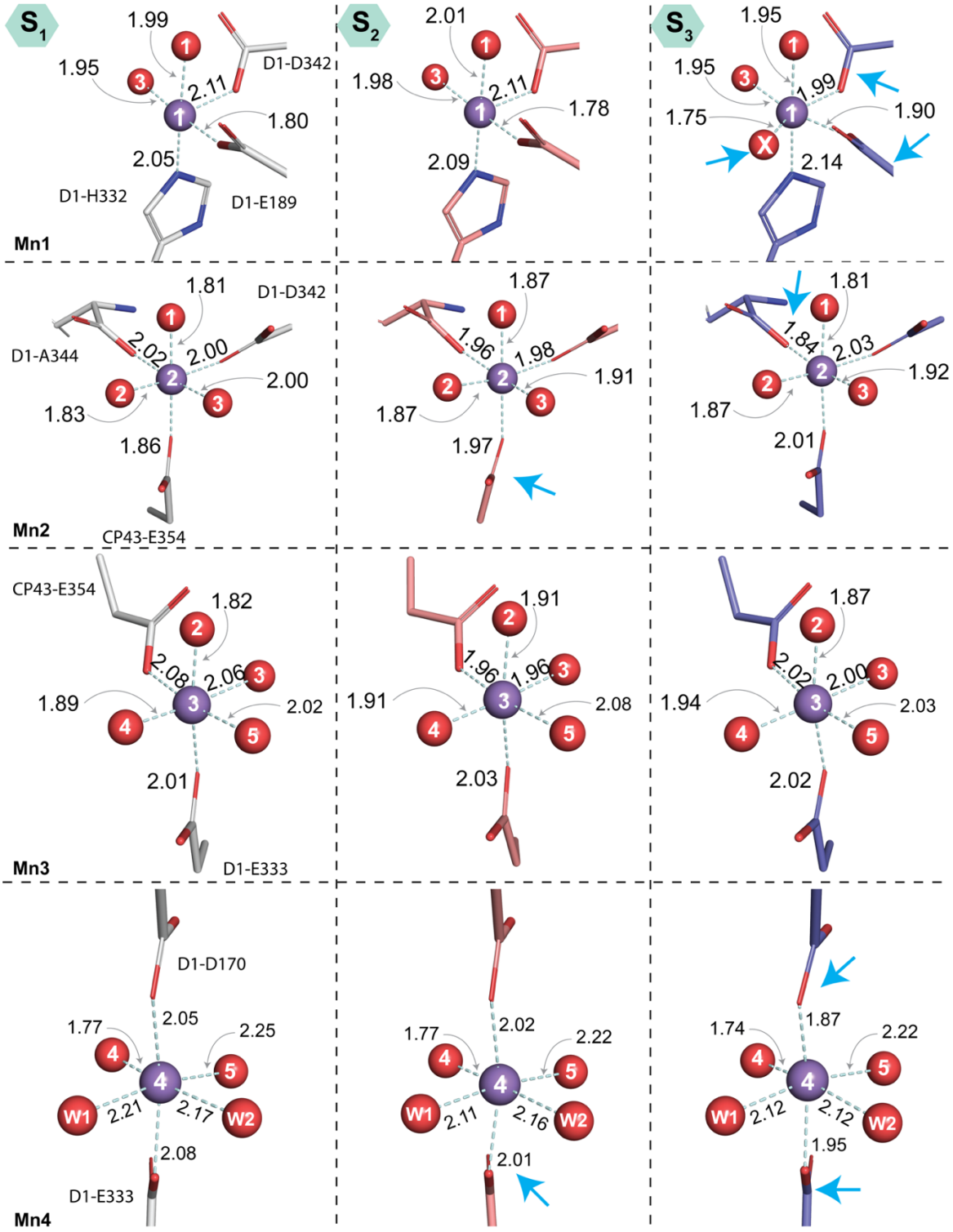
occupancy used in the modelling of the 2F state is obtained from omit maps and multiple spectroscopic techniques. Additional details of the plots in panel c are provided in Supplementary Methods. For panel (a), we do not show F_O - F_O map for 1F - 0F as a coordinate shift in monomer II was observed, leading to difference peaks on all 4 Mn centers. Source data for panel c is provided in the Source Data file



Supplementary Figure 4: Mn and Ca ligand environment in the S_1, S_2, S_3 states and coordination bond distances for monomer II: The OEC cluster and active site residues form coordination bonds with the Mn/Ca atoms. Shown are the coordination environment for (a) Mn4 (b) Mn1 and (c) Ca sites where the major structural changes occur. Mn, Ca and O atoms are shown as purple, green and red spheres respectively. The new inserted water in the S_3 state, O_X, is shown in yellow. Blue arrows denote the main changes observed. Significant distance contractions are highlighted with yellow shade and distance expansions with blue shade.



Supplementary Figure 5: Anomalous peak heights at the Mn centers in monomer I for the different S-states at 9.5 keV. Source data is provided in the Source Data file.

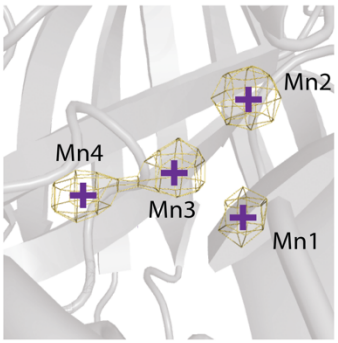


Supplementary Figure 6: Mn-ligand coordination environment in the different S-states in monomer I. Major changes are highlighted with blue arrows.

ANOMALOUS $f_{scatter} = f^0(s) + f'(\lambda) + if''(\lambda)$

1 Refine without OEC and evaluate anomalous signal

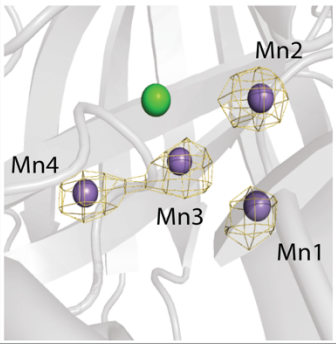
✚ = Mn centroid positions



Steps after refinement

-Place Mn in anomalous centroid postions

2 Unrestrained refinement of Mn coordinates and f′/f″.



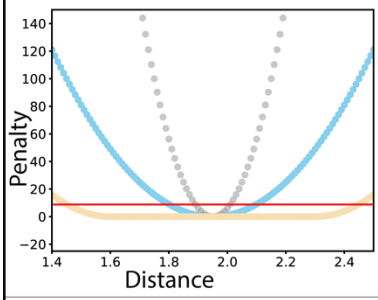
Steps after refinement

-Calculate Mn-Mn distances
-Place OEC oxygens (O1-O5, W1-W4) in residual F_O-F_C density

Box A
Restraints Validation

Method

- Refinement with Harmonic Restraints (HR)
- Additional refinement with control restraints (3b, 4b-d)
- Evaluate restraints

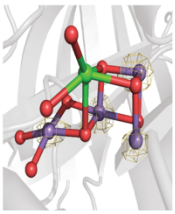


- HR, σ=0.05Å, our group
- FBHR, for validation
- HR, σ=0.02Å, CCP4 default
- Z = 2 boundary for restraints; criteria for severe penalty

Metrics for restraint evaluation

- RMSD with different restraints
- Restraint Z-scores
- Anomalous peak height of Mn
- F_O-F_C map flatness

3a Harmonic Mn-O restraints
No Mn-Mn restraints



Mean values from previous S_1 model σ=0.05 Å

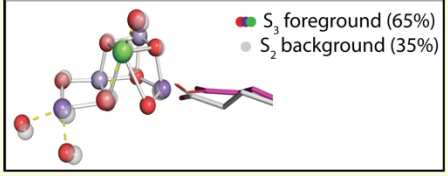
Steps after refinement

-Validate HR restraints (Box A)
-For 2F, proceed to step 3a'
-Move to Conventional Intensities step 4

3b Flat-bottom harmonic restraints
No Mn-Mn restraints

Bond distance ranges with no penalty:
Mn-O: 1.6-2.3 Å
Ca-O: 2.0-2.8 Å

3a′ For 2F only
Add O_x; Split the OEC+active site into two conformations



CONVENTIONAL

4a Harmonic Mn-O restraints
Anomalous Mn-Mn restraints

Mn-O restraints from 3a (σ=0.05 Å)
Mn-Mn distances restrained to anomalous distances (σ=0.01 Å)

Steps after refinement

-Validate HR restraints (Box A)
-Perform all subsequent model refinement

4b Flat-bottom harmonic restraints
No Mn-Mn restraints

Bond distance ranges with no penalty:
Mn-O: 1.6-2.3 Å
Ca-O: 2.0-2.8 Å

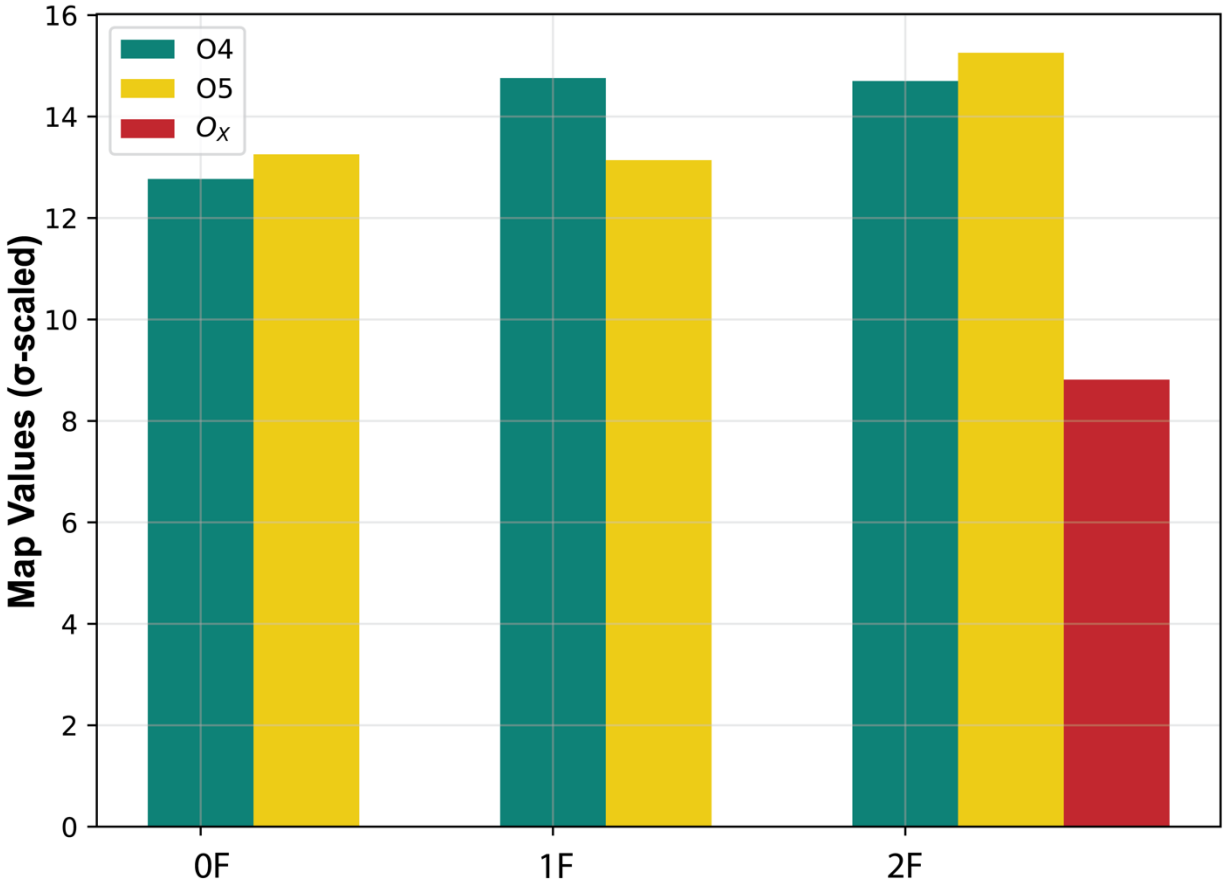
4c OEC restraints used in prior work

Mn-O and Mn-Mn restraints from spectroscopy (σ=0.05 Å)

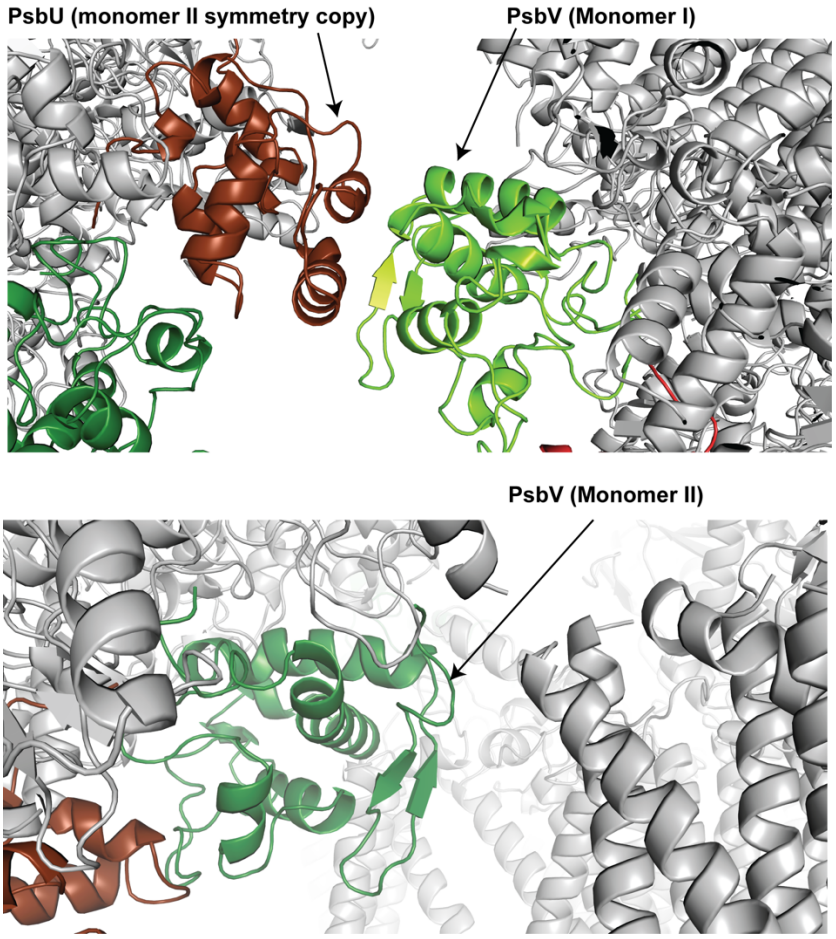
4d Default OEC restraints in refinement packages

Restraints from CCP4 mon lib (defaults) (σ=0.02 Å)

Supplementary Figure 7: Model building and refinement pipeline used in this paper. An approach combining information from anomalous and conventional intensities was developed to model the Photosystem II data. Earlier steps (1 - 3) used the anomalous data to model the Mn positions and build the OEC model. Subsequent steps (step 4 and beyond for final optimization) used conventional intensities, restraining the Mn to the anomalous density positions and determining the O positions with high precision. The OEC restraints were validated using multiple criteria listed in Box A and showed that the restraints did not bias the model. The boxes highlighted in green were chosen for the subsequent refinement steps. The boxes highlighted in grey are controls at the specified step of the modelling process. For final model building, optimization and deposition we used protocol 4a. Validation results are shown in Supplementary Table 3 and 4 and the restraints validation (Box A) is also discussed in the Methods section. A zoomed-in view of the residual plot in Box A is provided in Supplementary Figure 13.



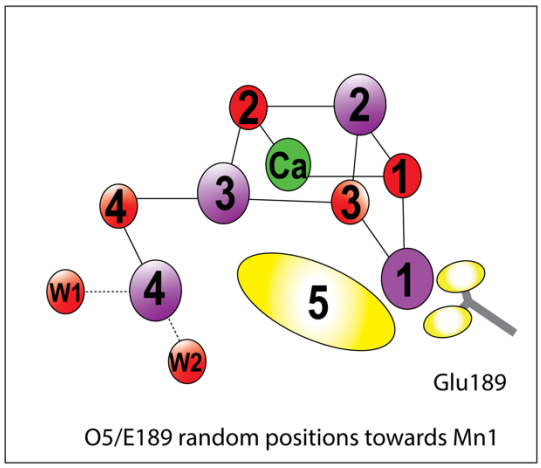
Supplementary Figure 8: Omit map peak heights for O4, O5 and O_X in monomer I for the different flash states from conventional diffraction data. Source data is provided in the Source Data file.



Supplementary Figure 9: PS II monomer naming convention based on crystal contacts. The PsbV chain in monomer I (light green, upper panel) makes a crystal contact with a symmetry copy of the PsbU chain (brown) in monomer II. No such crystal contact is observed for PsbV in monomer II (dark green, lower panel).

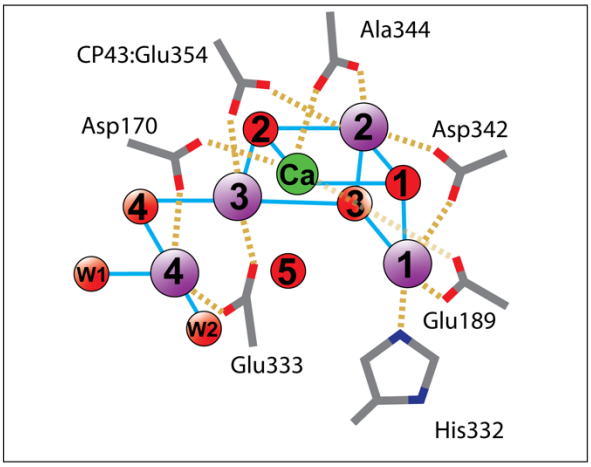
(a)

1. Start with 1-component model with no O_x
2. Move O5/E189 randomly towards Mn1
3. Shake all coordinates/B-factors
4. Create 122 starting conformations



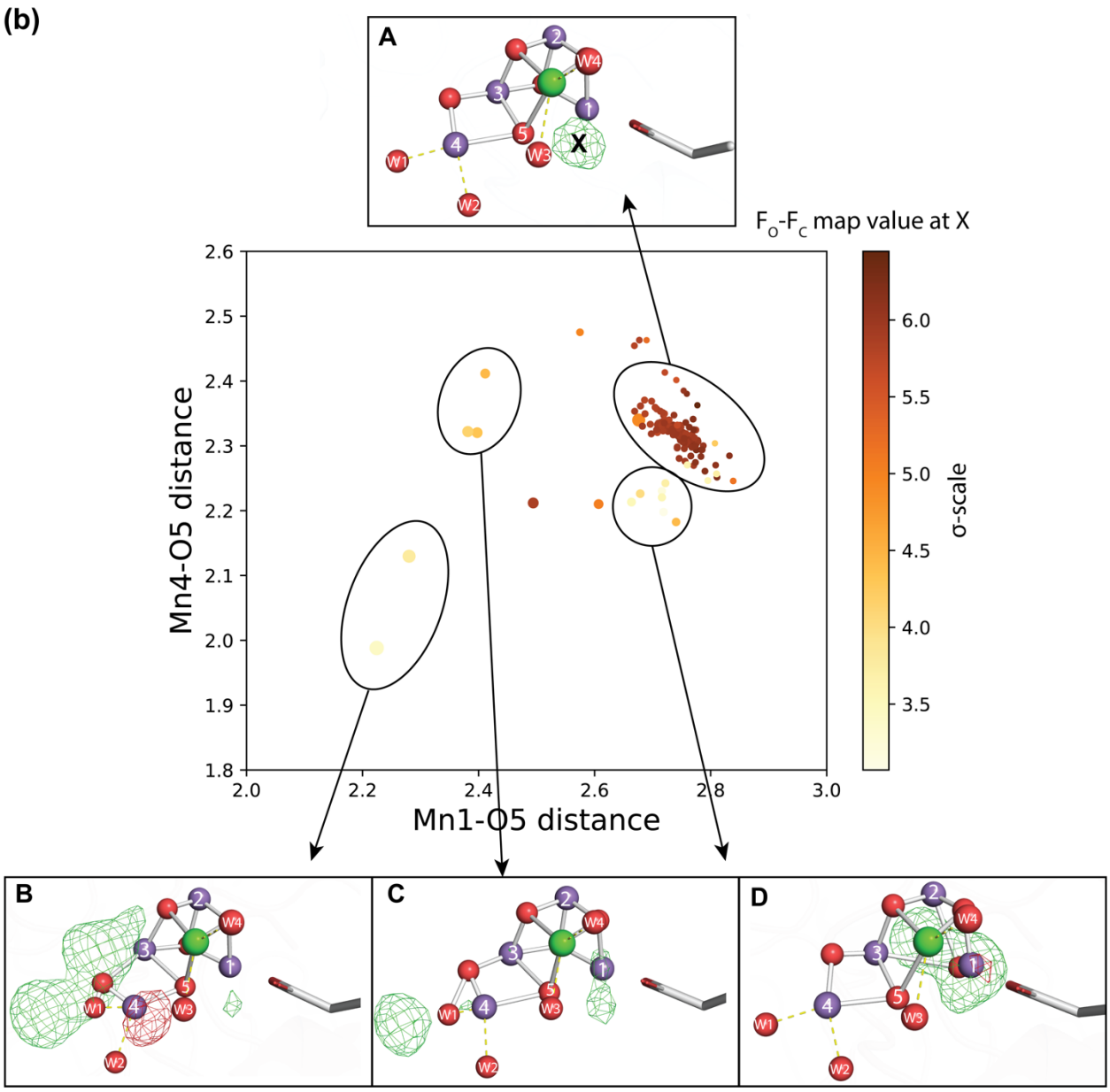
O5/E189 random positions towards Mn1

5. FBHR restraints for OEC and loose Mn-protein coordination
6. No restraint on O5, free to move
7. No OEC angles or any topology specified



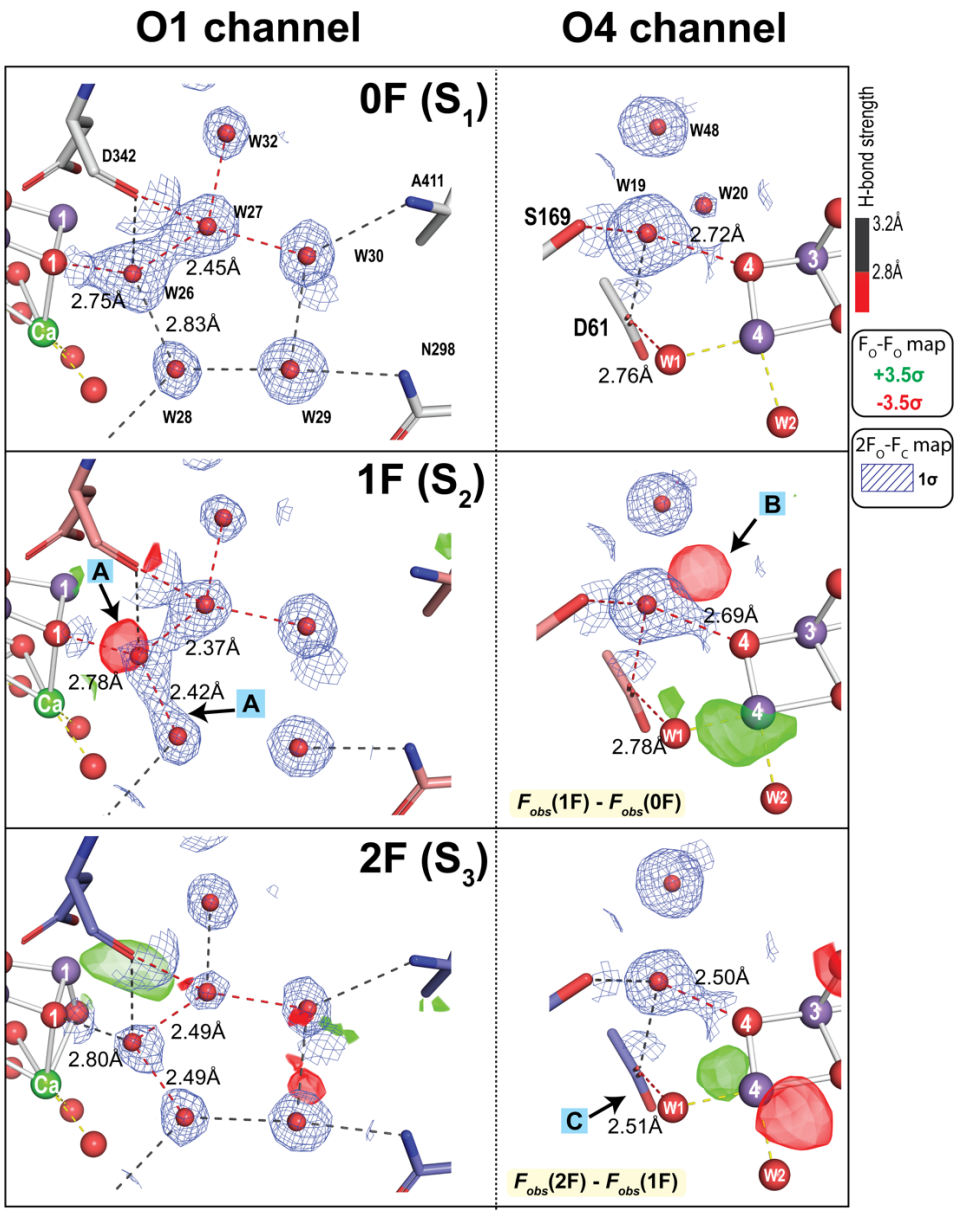
FBHR restraints for Mn-O bonds
No penalty between 1.6-2.3Å

Mn-protein coordination bond
σ = 0.50Å

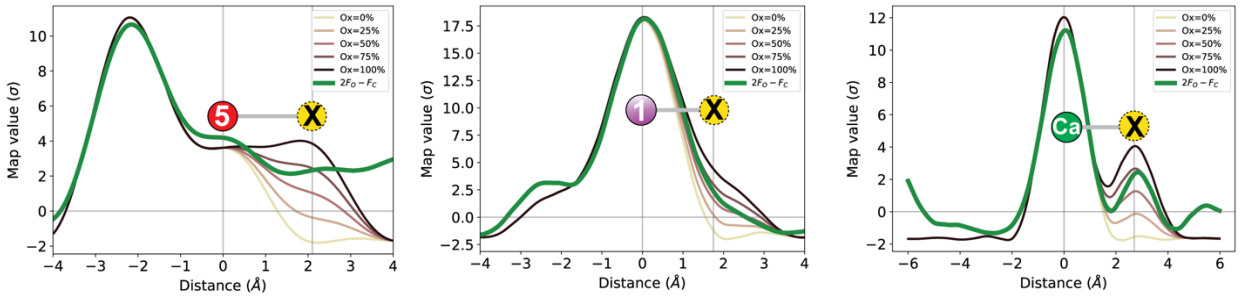


Supplementary Figure 10: Single-conformer model for the 2F dataset with no O_X. Multiple independent refinements (143 trials) of the 2F data with single-conformer models were run. (a) Steps to create the starting conformers and the OEC restraints used during refinement. The O5 and E189 positions were pushed closer to the F_O - F_C peak (marked as X in panel b) instead of modelling an extra ligand (O_X) to create different starting conditions. No Mn-O5 restraints were applied. FBHR restraints and very loose Mn-protein coordination were used, allowing maximum freedom during refinement for the OEC with no topology enforced. The full Ca environment is not shown for clarity (b) Refinement results from the trials. Each dot in the middle plot represents

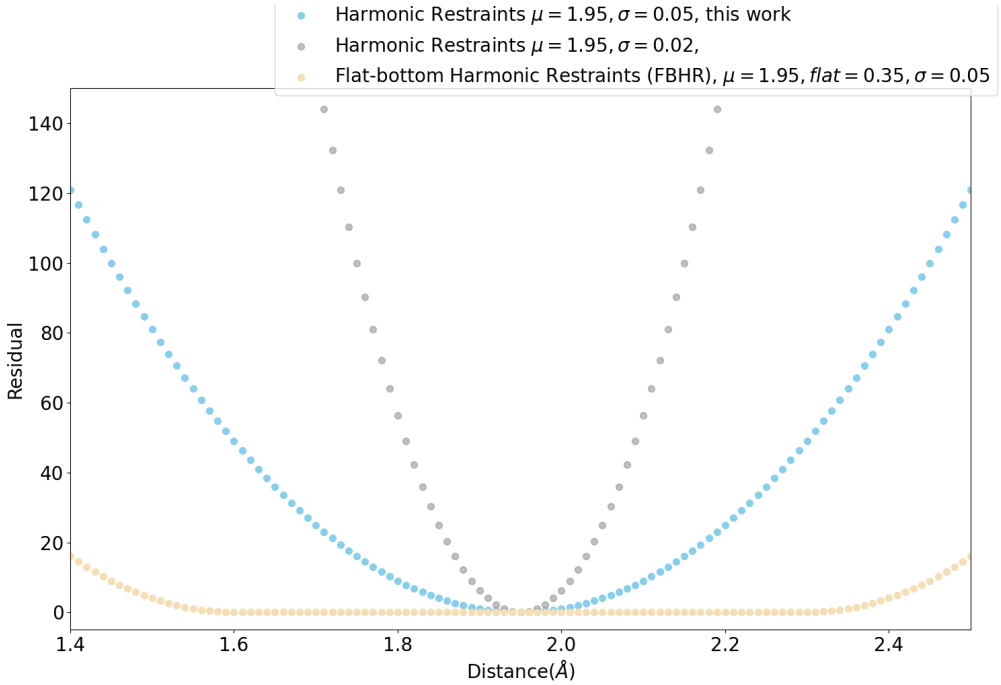
a refinement result with the color bar representing the F_O-F_C peak height at the X position and the size of the dots are proportional to the deviation from the Mn1-Mn4 distance in the S_3 state (5.05 Å). The F_O - F_C maps shown are contoured at 4σ. In the majority cluster (A, top panel), the O5 stays coordinated with Mn4 with a residual 6σ F_O-F_C peak at the X position (top figure). The E189 conformation refined to the S_3 state conformation in all cases. Representative examples from three other clusters are shown in B, C and D. In all three cases (and in general in the 3 clusters), refinement led to a severely distorted OEC (seen from unphysical W1-O4 distances in cases B, C or Mn1-O3 distances in D). Large F_O - F_C blobs are also seen around the OEC in B, C, D indicating refinement is stuck in a local minimum and is unable to suppress the difference features with a physically plausible model. The analysis demonstrates that a 2-conformer model with an additional ligand (O_X) bridging Mn1 and Ca is necessary to explain the 2F electron density data with O5 staying coordinated to the Mn4 site. Source data for panel b is provided in the Source Data file.



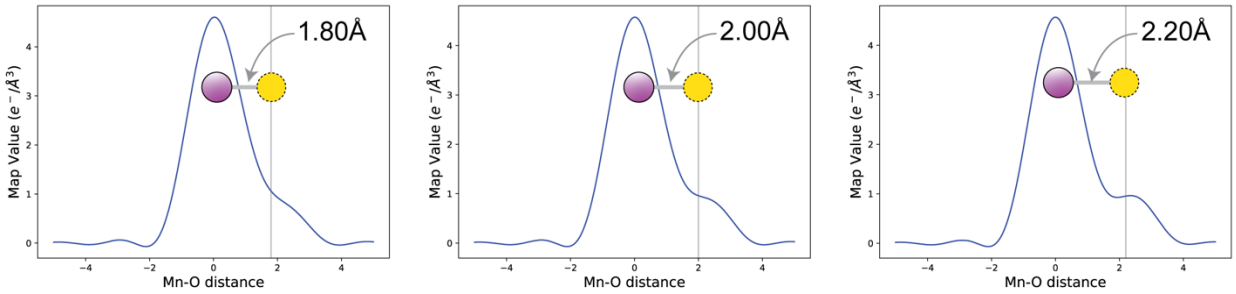
Supplementary Figure 11: Changes in the water and hydrogen-bond network near the OEC. Several notable changes are observed in the O1 channel (left panel) and O4 channel (right panel) in the S_1, S_2 and S_3 states as highlighted in the figure (A) Change in hydrogen-bond distances in the water wheel and increased mobility of the water wheel, especially W26, W27, W28 in the S_2 state. (B) Disappearance or disorder of W20 in the O4 channel in the S_2 state (C) Strong hydrogen-bond interaction between D61-W1 in the S_3 state. Shown are $2F_O$-F_C maps contoured at 1.0σ in blue mesh and F_O-F_O maps contoured at +/- 3.5σ in green/red with the S_1, S_2 and S_3 state models overlaid in the respective panels.



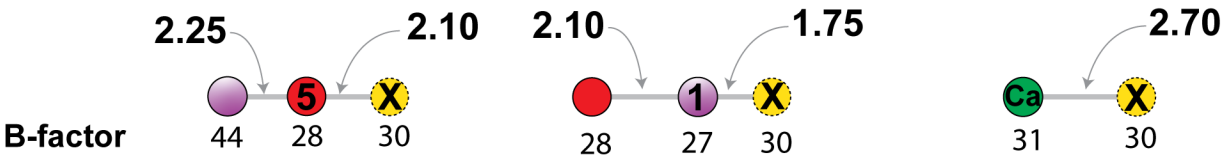
Supplementary Figure 12: F_{calc} map traces along bond vectors of simplified 2 or 3-atom models and comparison with the experimental 2F data. The maps were calculated in the 2F unit cell using only the subset of relevant atoms (described in methods) for the O5-O_X, Mn1-O_X, Ca-O_X vectors and the same set of miller indices as the experimental map. O_X occupancies were varied from 0-100%. The results are similar to the observations from the theoretical Fourier maps shown in figure 3C and support the modelling of the ligand O_X in the 2F data. Source data is provided in the Source Data file.



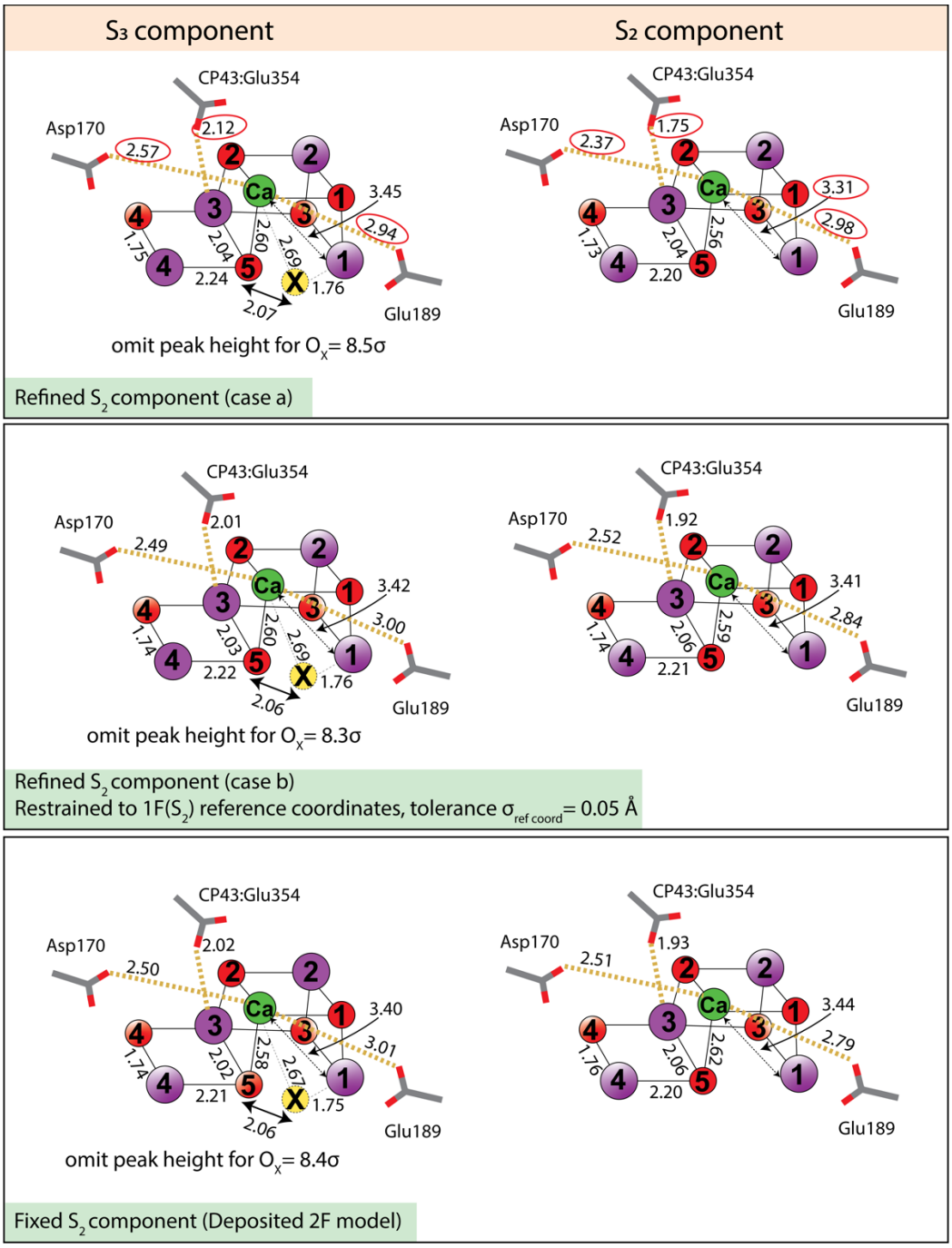
Supplementary Figure 13: Residual plot comparing different restraint types including effect of different σ (expected standard deviation of the bond distances). For this example, we used a mean bond distance (μ) = 1.95 Å. In this work and our previous papers, we used harmonic restraints (HR) similar to the blue profile with $\sigma = 0.05$ Å. Standard crystallographic packages provide default restraints for OEX/OEY ligand (Mn_4Ca cluster) similar to the grey profile with $\sigma = 0.02$ Å . In order to check any possibilities of bias in our HR restraints, we also ran separate refinements with the FBHR restraints (wheat profile), details of which are provided in the Methods. Source data is provided in the Source Data file.



Supplementary Figure 14: Theoretical error-free Mn-O electron density traces with different Mn-O distances at the resolution of the 2F data (1.93 Å). These traces show the expected level of oxygen scattering signal in the vicinity of heavy metal atoms (Mn). The Mn and O occupancies were 100% and 75% respectively. It is noteworthy for the Mn-O distance = 1.80 Å plot that the bulge in the profile is shifted away from the true oxygen position. This highlights how the theoretical profiles are a useful way to correctly interpret experimental maps. Source data is provided in the Source Data file.

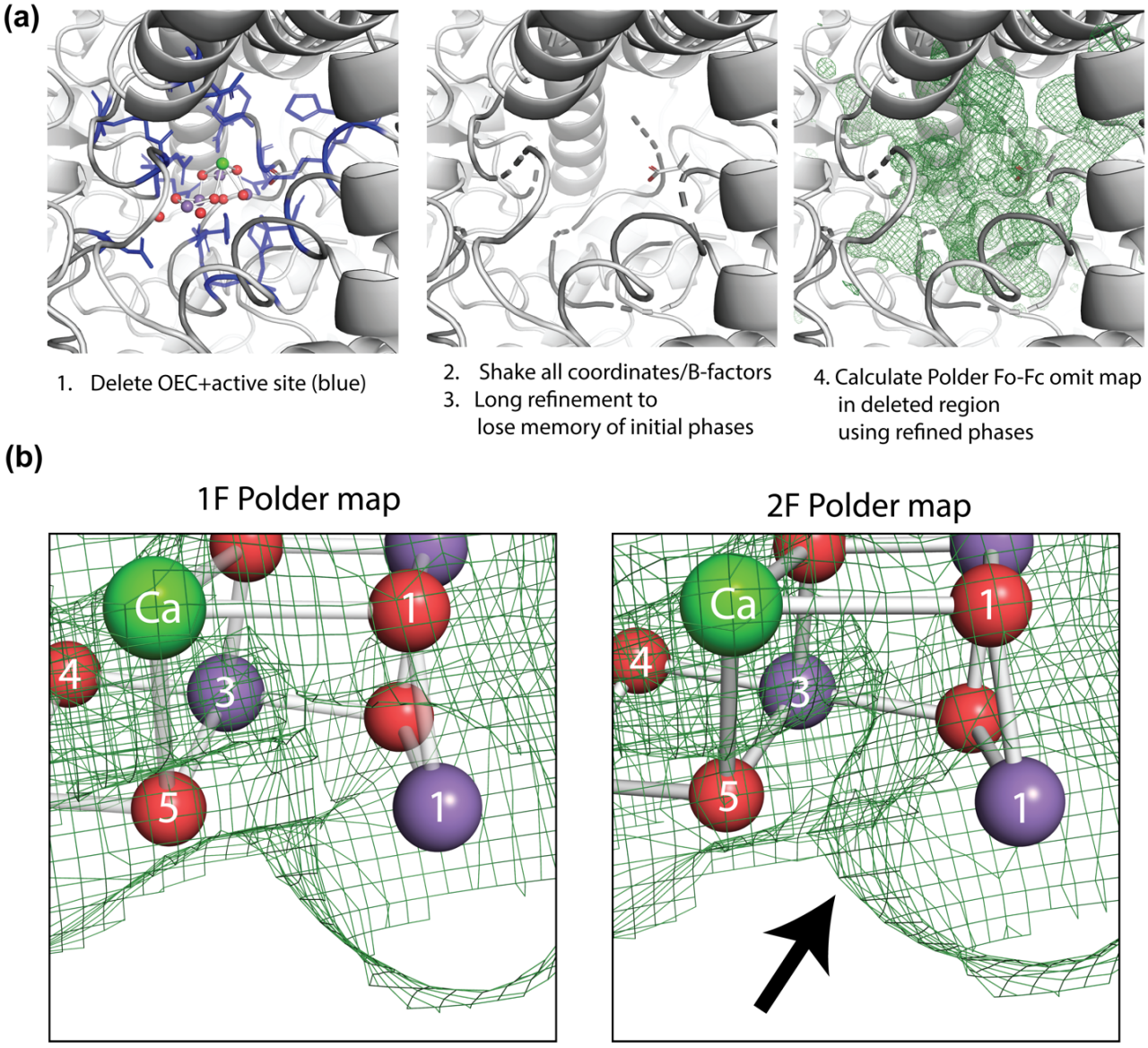


Supplementary Figure 15: Parameters used for theoretical electron density profiles for 2- or 3-atom models. The bond distance and B-factor parameters were used to obtain the electron density profiles shown in Figure 3c in the main text and its comparison with the $2F_O$-F_C map profiles.



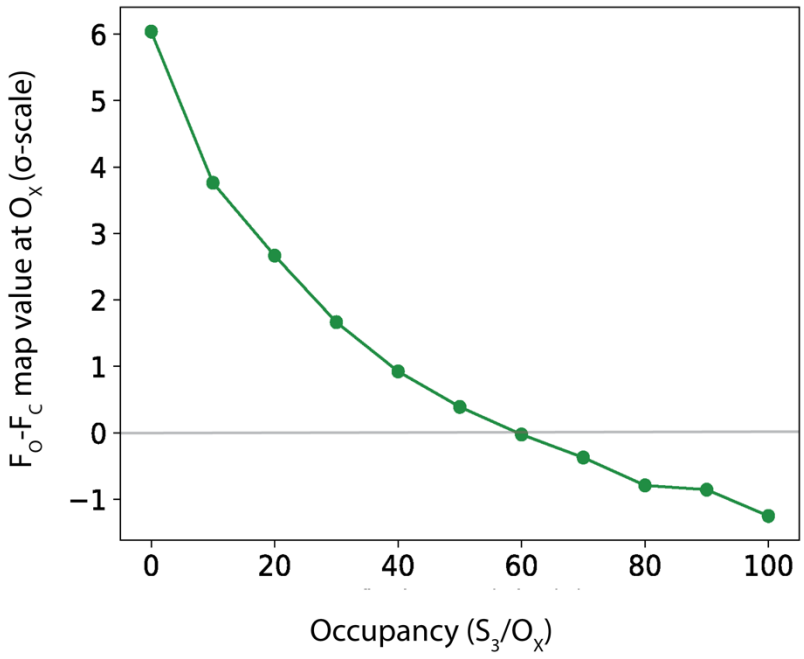
Supplementary Figure 16: Effect of refining S_2 component in the 2F model. Representative bond distances and O_X geometry under different refinement conditions. *Top Panel*: Case (a) as described in Supplementary Discussion 5 where the S_2 component was allowed to refine. *Middle Panel*: Case (b) where limited coordinate refinement was allowed for the S_2 component. *Bottom panel*: S_2 conformation is kept constrained to the 1F(S_2) structure. This 2F model is described in this paper. Significant changes for certain bonds are seen in case (a), highlighted with red circles,

especially for the S_2 component, underscoring issues in refining S_2 and S_3 simultaneously without additional information. The O_X electron density and geometry is similar in all cases.



Supplementary Figure 17: Sensitivity of O_X signal to active site phases. (a) Steps to calculate Polder F_O-F_C omit maps from phases with no memory of the active site. The modified phases were obtained by deleting the active site residues, waters and the OEC and subsequent randomization and long refinement. A refinement of 100 cycles was carried out (normally 15-30 cycles) to obtain phases that do not have memory of the OEC and active site region. Details of the steps are described in the Map Calculations section in Supplementary Information Methods (b) Resulting maps are shown at a 5σ level in the 1F and 2F state with the final 1F/2F models overlaid. The O_X is not shown in the 2F model for visual clarity. A distinct additional density in the form of a bulge

is visible in the 2F map in the vicinity of the Mn1 location (black arrow). Such a feature is not present in the 1F control. Map features for all other oxygen ligands including O5 are similar between the 2 states. This additional density is evidence for the extra oxygen ligand, O_X. The analysis highlights the robustness of the O_X signal to phases from the surrounding active site.



Supplementary Figure 18: Estimating S_3/O_X population. A set of two-component 2F models with the S_3 component (containing O_X) varying between 0 - 100% in steps of 10% were refined. The S_2/S_3 population in each model was kept fixed and only coordinates/B-factors of S_3 were refined, similar to the deposited model. The residual F_O-F_C map value at the O_X position was evaluated. Since an optimal model would be able to minimize F_O-F_C differences at the O_X position, a population 60-70% is most probable from this analysis. This is consistent with other approaches and the final O_X population of 65% used in the paper. Source data is provided in the Source Data file.

Supplementary Tables

Supplementary Table 1: Data collection, merging and refinement statistics (0F, 1F, 2F) for the 9.5keV data

Data set[⊥]	0F-Anom.	0F-Conv.	1F-Anom.	1F-Conv.	2F-Anom.	2F-Conv.
PDB ID	9Z33	9Z68	9Z34	9Z69	9Z35	9Z6A
Resolution range refined (Å)	35.14 – 1.97	35.14 – 1.97	37.14 – 1.98	37.14 – 1.98	33.83 – 1.93	33.83 – 1.93
Resolution range upper bin (Å)	2.00 – 1.97	2.00 – 1.97	2.01 – 1.98	2.01 – 1.98	1.96 – 1.93	1.96 – 1.93
Wavelength (Å)	1.317	1.317	1.3062	1.3062	1.3056	1.3056
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell parameters (Å)	a=117.5 b=222.6 c=308.9	a=117.5 b=222.6 c=308.9	a=117.5 b=222.8 c=309.1	a=117.5 b=222.8 c=309.1	a=117.5 b=222.8 c=308.9	a=117.5 b=222.8 c=308.9
Lattices merged	24239	24239	10764	10764	26194	26194
Unique reflections	1103017	566181	1089133	559068	1174348	602738
(upper bin)	52636	27995	52094	27688	55696	29869
Completeness (%)	99.69	99.95	99.70	99.96	99.65	99.96
(upper bin)	95.16	99.62	95.27	99.69	94.38	99.66
$CC_{1/2}$	0.993	0.996	0.987	0.994	0.991	0.996
(upper bin)	0.072	0.087	0.067	0.115	0.058	0.074
$I/\sigma_{MM24}(I)^{\dagger}$	4.383	6.33	3.726	5.174	4.168	5.791
(upper bin)	0.303	0.44	0.355	0.493	0.306	0.428
Wilson B-factor ($Å^2$)	35.64	35.59	35.08	35.15	33.82	33.83
R-factor (%)	19.42	18.79	18.81	17.27	18.90	17.48
R-free (%)	23.03	22.60	22.39	21.30	22.31	21.23
Number of atoms	52714	104125	52680	104568	54713	105277
Number non-hydrogen atoms	52714	52043	52680	52483	54713	54720
Ligands atoms	8949	8941	8949	8941	8979	8963
Waters	2168	1505	2134	1945	2735	2758
Protein residues	5299	5299	5299	5299	5299	5299
RMS (bonds)	0.009	0.012	0.009	0.010	0.010	0.008
RMS (angles)	1.179	1.383	1.179	1.382	1.176	1.12
Ramachandran favored (%)	97.64	97.30	97.72	97.64	97.85	97.77
Ramachandran outliers (%)	0.25	0.21	0.19	0.15	0.19	0.15

Rotamer favored (%)	91.53	91.46	92.46	92.04	92.28	93.17
Rotamer outliers (%)	1.78	1.02	1.32	0.86	1.52	2.80
Rama-Z score	-0.75	-1.21	-0.65	-0.97	-0.56	-0.81
C_β deviation (%)	0.00	0.00	0.00	0.00	0.00	0.00
Clashscore	4.78	3.49	3.98	3.02	3.96	2.83
Average B-factor (Å^2)	42.59	43.72	40.31	42.26	39.56	41.75

$^{\perp}$ The raw diffraction data for the anomalous and conventional datasets for each flash state is the same with the merging of structure factor intensities done differently. See Methods for details.
† As defined in Supplementary reference 3

Supplementary Table 2: Data collection, merging and refinement statistics (0F, 2F) for the 7.0keV data

Data set	0F-Anom.	2F-Anom.
PDB ID	9Z36	9Z37
Resolution range refined (Å)	37.99 – 2.16	37.85 – 2.11
Resolution range upper bin (Å)	2.20 – 2.16	2.15 – 2.11
Wavelength (Å)	1.7715	1.7715
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit cell parameters (Å)	a=117.6 b=222.9 c=309.5	a=117.6 b=222.8 c=309.5
Lattices merged	9357	12257
Unique reflections	840035	901117
(upper bin)	40734	43481
Completeness (%)	99.69	99.66
(upper bin)	96.45	96.08
$CC_{1/2}$	0.975	0.976
(upper bin)	0.081	0.071
$I/\sigma_{MM24}(I)^{\dagger}$	3.237	3.365
(upper bin)	0.435	0.375
Wilson B-factor (Å^2)	38.39	37.73
R-factor	19.77	19.36
R-free	23.98	23.40
Number of atoms	52923	54355
Number non-hydrogen atoms	52923	54355
Ligands	8949	8979
Waters	2377	2377

Protein residues	5299	5299
RMS (bonds)	0.009	0.010
RMS (angles)	1.184	1.188
Ramachandran favored (%)	97.53	97.56
Ramachandran outliers (%)	0.17	0.17
Rotamer favored (%)	90.88	90.19
Rotamer outliers (%)	1.67	1.63
RamaZ score	-0.95	-0.93
C_β deviation (%)	0.00	0.00
Clashscore	4.98	5.22
Average B-factor (Å^2)	43.52	43.00

Supplementary Table 3: **Effect of using different restraints (steps 3a, 3b in Supplementary Figure 7) on the geometry of the Mn_4Ca cluster in the 0F, 1F and 2F states**. Anomalous intensities were used as input for these steps. The 2F state data shown is from a 1-component model with no O_X modelled. Based on the validation results, output from step 3a were used to do additional model building. All distances are in units of Angstroms.

	0F		1F		2F (1-component model)	
	Harmonic restraints (3a)	Flat-bottom Harmonic Restraints (FBHR) (3b)	Harmonic restraints (3a)	Flat-bottom Harmonic Restraints (FBHR) (3b)	Harmonic restraints (3a)	Flat-bottom Harmonic Restraints (FBHR) (3b)
MN1:MN3	3.22	3.19	3.24	3.24	3.34	3.3
MN1:MN4	4.82	4.81	4.84	4.86	4.98	4.94
MN3:MN4	2.77	2.81	2.72	2.72	2.76	2.73
MN1:MN2	2.80	2.8	2.79	2.78	2.83	2.81
MN2:MN3	2.82	2.8	2.83	2.83	2.82	2.82
CA1:MN1	3.44	3.43	3.42	3.45	3.37	3.35
CA1:MN3	3.54	3.55	3.52	3.54	3.53	3.52
CA1:MN2	3.44	3.42	3.49	3.51	3.4	3.42
MN1:O1	2.00	2	2.00	1.97	1.99	2.02
MN1:O3	1.94	1.99	1.95	2.03	1.94	2.05
MN2:O1	1.77	1.75	1.83	1.85	1.81	1.79
MN2:O2	1.85	1.79	1.86	1.82	1.85	1.86
MN2:O3	1.96	2.04	1.91	1.95	1.92	1.97
MN3:O2	1.84	1.84	1.91	2	1.85	1.85
MN3:O4	1.85	1.65	1.87	1.78	1.88	1.83
MN3:O5	2.02	2.06	2.05	2.12	2.08	2.21
MN4:O4	1.79	1.85	1.76	1.82	1.74	1.74
MN4:O5	2.22	2.26	2.22	2.22	2.23	2.33
MN4:W1	2.21	2.28	2.13	2.09	2.08	2.05
MN4:W2	2.15	2.18	2.13	2.12	2.11	2.11
CA1:O2	2.49	2.54	2.50	2.52	2.52	2.57
CA1:O1	2.34	2.37	2.33	2.39	2.34	2.38
CA1:O5	2.58	2.6	2.59	2.66	2.59	2.5
CA1:W3	2.51	2.5	2.52	2.49	2.52	2.45
CA1:W4	2.32	2.27	2.30	2.22	2.31	2.27

Supplementary Table 4: **Effect of using different restraints (steps 4a, 4b, 4c in Supplementary Figure 7) on the geometry of the Mn_4Ca cluster in the 0F, 1F and 2F states**. Conventional intensities were used as input for these steps. The 2F state data shown is from a 2-component model. Protocol 4a was used for final model optimization shown in this paper and PDB models 9Z68 (0F), 9Z69 (1F) and 9Z6A (2F). All distances are in units of Angstroms.

	0F			1F			2F (2-component)		
	Mn-O restraints + Mn-Mn restraints from anomalous (4a)	Mn-O restraints using FBHR (4b)	OEC restraints from PNAS 2020 (4c)	Mn-O restraints + Mn-Mn restraints from anomalous (4a)	Mn-O restraints using FBHR (4b)	OEC restraints from PNAS 2020 (4c)	Mn-O restraints + Mn-Mn restraints from anomalous (4a)	Mn-O restraints using FBHR (4b)	OEC restraints from PNAS 2020 (4c)
MN1:MN3	3.22	3.22	3.24	3.25	3.26	3.26	3.41	3.41	3.38
MN1:MN4	4.83	4.84	4.81	4.86	4.87	4.87	5.14	5.14	5.08
MN3:MN4	2.78	2.78	2.75	2.72	2.73	2.75	2.76	2.76	2.76
MN1:MN2	2.8	2.8	2.78	2.79	2.78	2.77	2.8	2.8	2.79
MN2:MN3	2.82	2.82	2.83	2.84	2.84	2.83	2.83	2.83	2.85
CA1:MN1	3.44	3.43	3.42	3.45	3.46	3.44	3.40	3.41	3.39
CA1:MN3	3.59	3.6	3.59	3.57	3.58	3.55	3.56	3.53	3.55
CA1:MN2	3.44	3.42	3.43	3.52	3.53	3.51	3.4	3.39	3.35
MN1:O1	1.99	1.98	2.02	2.01	1.96	1.99	1.95	1.97	1.96
MN1:O3	1.95	1.98	1.94	1.98	2.04	1.95	1.95	2.13	1.92
MN2:O1	1.81	1.78	1.79	1.87	1.93	1.85	1.81	1.8	1.82
MN2:O2	1.83	1.82	1.83	1.87	1.83	1.85	1.87	1.84	1.86
MN2:O3	2.00	2.08	1.98	1.91	1.91	1.91	1.92	2.01	1.87
MN3:O2	1.82	1.83	1.84	1.91	1.99	1.91	1.87	1.84	1.86
MN3:O3	2.06	2.20	2.04	1.96	1.94	1.96	2.00	1.93	1.99
MN3:O4	1.89	1.81	1.88	1.91	1.91	1.89	1.94	1.83	1.92
MN3:O5	2.02	1.96	2.02	2.08	2.13	2.05	2.03	2.13	2.03
MN4:O4	1.77	1.84	1.77	1.77	1.8	1.75	1.74	1.69	1.75
MN4:O5	2.25	2.23	2.22	2.22	2.2	2.19	2.22	2.3	2.23
MN4:W1	2.21	2.29	2.2	2.11	2.1	2.13	2.12	1.99	2.09
MN4:W2	2.17	2.22	2.16	2.16	2.19	2.14	2.12	2.1	2.11
CA1:O2	2.52	2.53	2.51	2.52	2.56	2.53	2.52	2.64	2.52
CA1:O1	2.38	2.42	2.36	2.37	2.39	2.35	2.39	2.4	2.28
CA1:O5	2.62	2.67	2.6	2.62	2.69	2.62	2.59	2.48	2.58
CA1:W3	2.55	2.56	2.54	2.52	2.51	2.53	2.52	2.43	2.53
CA1:W4	2.32	2.37	2.36	2.33	2.3	2.3	2.34	2.23	2.32
MN1:O_X							1.75	1.59	1.77
CA1:O_X							2.68	2.81	2.63
O5:O_X							2.06	2.04	2.08

Supplementary Table 5: **Comparison of distances between centroids for Mn from anomalous diffraction map, 2Fo-Fc map and the final Mn-Mn distances from refined models (both 9.5 keV and 7 keV).** The anomalous diffraction map and 2Fo-Fc map are from step 1 refinement when no Mn is present in the model. The centroids were calculated within a 1 Å radius of the final Mn coordinates (taken from step 4). All distances are in units of Angstroms.

9.5 keV data									
	0F			1F			2F		
	Anomalous Centroid	2Fo-Fc centroid	Refined model (S_1)	Anomalous Centroid	2Fo-Fc centroid	Refined model (S_2)	Anomalous Centroid	2Fo-Fc centroid	Refined model (S_3)
Mn1-Mn4	4.84	4.83	4.83	4.93	4.87	4.86	5.14	5.11	5.14
Mn1-Mn3	3.26	3.27	3.22	3.26	3.30	3.25	3.41	3.43	3.41
Mn3-Mn4	2.85	2.72	2.78	2.75	2.69	2.72	2.77	2.72	2.76
Mn1-Mn2	2.75	2.82	2.80	2.84	2.80	2.79	2.81	2.82	2.80
Mn2-Mn3	2.74	2.83	2.82	2.86	2.85	2.84	2.85	2.84	2.83
7 keV data									
	0F						2F		
	Anomalous Centroid	2Fo-Fc centroid	Refined model (S_1)				Anomalous Centroid	2Fo-Fc centroid	Refined model (S_3)
Mn1-Mn4	4.82	4.86	4.82				5.07	5.08	5.06
Mn1-Mn3	3.21	3.25	3.19				3.37	3.39	3.36
Mn3-Mn4	2.78	2.76	2.79				2.74	2.69	2.72
Mn1-Mn2	2.89	2.85	2.84				2.84	2.81	2.83
Mn2-Mn3	2.83	2.89	2.85				2.89	2.92	2.88

Supplementary Table 6: **Improved resolution and Mn position modelling leads to more precise Mn-O distances:** The Table compares the END/RAPID error estimates from the current 0F conventional data (1.97 Å) with a previously published 0F data (PDB 6W1O, 2.08 Å). Note that the bond length error estimates are an upper bound to the true error since we use a much larger perturbation (F_O-F_C differences) that expected $\sigma(F_O)$ to simulate the experimental errors in structure factors. See the END/RAPID section in Methods for more details. All distances are in units of Angstroms.

	Current 0F data (1.97 Å)		PDB 6W1O (2.08 Å)	
Pair	PDB distance	Error upper bound	PDB distance	Error upper bound
MN1-MN4	4.83	0.04	4.81	0.07
MN1-MN3	3.22	0.03	3.24	0.08
MN3-MN4	2.78	0.03	2.67	0.11
MN1-MN2	2.80	0.03	2.77	0.05
MN2-MN3	2.82	0.03	2.9	0.1
MN1-O1	1.99	0.06	2.01	0.09
MN1-O3	1.95	0.15	1.92	0.13
MN1-His332	2.05	0.05	2.13	0.06
MN1-Glu189	1.80	0.05	1.74	0.06
MN1-Asp342	2.11	0.05	2.15	0.05
MN2-O1	1.81	0.08	1.76	0.09
MN2-O2	1.83	0.07	1.84	0.12
MN2-O3	2.00	0.08	1.99	0.1
MN2-Asp342	2.00	0.05	1.99	0.06
MN2-Glu354	1.86	0.05	1.86	0.06
MN2-Ala344	2.02	0.05	1.87	0.07
MN3-O2	1.82	0.06	1.86	0.1
MN3-O3	2.06	0.14	1.93	0.14
MN3-O4	1.89	0.1	1.85	0.21
MN3-O5	2.02	0.09	1.98	0.12
MN3-Glu333	2.01	0.06	1.98	0.06
MN3-Glu354	2.08	0.08	1.95	0.08

MN4-O4	1.77	0.08	1.75	0.15
MN4-O5	2.25	0.11	2.13	0.16
MN4-W1	2.21	0.08	2.12	0.11
MN4-W2	2.17	0.08	2.15	0.11
MN4-Asp170	2.05	0.04	1.96	0.06
MN4-Glu333	2.08	0.05	2.04	0.06
CA1-O1	2.38	0.08	2.34	0.09
CA1-O2	2.52	0.07	2.45	0.1
CA1-O5	2.62	0.1	2.55	0.11
CA1-W3	2.55	0.1	2.56	0.11
CA1-W4	2.32	0.09	2.27	0.1
CA1-Asp170	2.40	0.06	2.46	0.07
CA1-Glu189	2.92	0.06	2.91	0.09
CA1-Ala344	2.54	0.06	2.62	0.09

Supplementary Table 7: Error estimates on the different OEC coordination bond distances using the END/RAPID procedure with conventional intensities. All distances are in units of Angstroms.

	0F		1F		2F	
	Distance	Error	Distance	Error	Distance	Error
MN1-MN4	4.83	0.04	4.86	0.04	5.14	0.06
MN1-MN3	3.22	0.03	3.25	0.03	3.41	0.03
MN3-MN4	2.78	0.03	2.72	0.03	2.76	0.03
MN1-MN2	2.80	0.03	2.79	0.02	2.80	0.02
MN2-MN3	2.82	0.03	2.84	0.03	2.83	0.03
MN1-O1	1.99	0.06	2.01	0.07	1.95	0.06
MN1-O3	1.95	0.15	1.98	0.06	1.95	0.12
MN1-HIS332	2.05	0.05	2.09	0.04	2.14	0.05
MN1-GLU189	1.80	0.05	1.78	0.04	1.90	0.06
MN1-ASP342	2.11	0.05	2.11	0.04	1.99	0.06
MN2-O1	1.81	0.08	1.87	0.07	1.81	0.06
MN2-O2	1.83	0.07	1.87	0.09	1.87	0.1
MN2-O3	2.00	0.08	1.91	0.07	1.92	0.09
MN2-ASP342	2.00	0.05	1.98	0.04	2.03	0.05
MN2-GLU354	1.86	0.05	1.97	0.05	2.01	0.05
MN2-ALA344	2.02	0.05	1.96	0.05	1.84	0.07
MN3-O2	1.82	0.06	1.91	0.07	1.87	0.07
MN3-O3	2.06	0.13	1.96	0.1	2.00	0.1
MN3-O5	2.02	0.09	2.08	0.09	2.03	0.11
MN3-GLU333	2.01	0.06	2.03	0.04	2.02	0.06
MN3-O4	1.89	0.1	1.91	0.1	1.94	0.1
MN3-GLU354	2.08	0.08	1.96	0.05	2.02	0.06
MN4-O4	1.77	0.08	1.77	0.08	1.74	0.08
MN4-O5	2.25	0.11	2.22	0.11	2.22	0.1
MN4-W1	2.21	0.08	2.11	0.07	2.12	0.1
MN4-W2	2.17	0.08	2.16	0.07	2.12	0.09
MN4-ASP170	2.05	0.04	2.02	0.04	1.87	0.05
MN4-GLU333	2.08	0.05	2.01	0.04	1.95	0.06

CA1-O1	2.38	0.08	2.37	0.07	2.39	0.08
CA1-O2	2.52	0.07	2.52	0.07	2.52	0.12
CA1-O5	2.62	0.1	2.62	0.08	2.59	0.1
CA1-W3	2.55	0.1	2.52	0.06	2.52	0.1
CA1-W4	2.32	0.09	2.33	0.06	2.34	0.1
CA1-ASP170	2.40	0.06	2.48	0.06	2.50	0.07
CA1-GLU189	2.92	0.06	2.79	0.04	3.01	0.08
CA1-ALA344	2.54	0.06	2.54	0.05	2.56	0.06
MN1-O_X					1.75	0.10
CA1-O_X					2.68	0.09
O5-O_X					2.06	0.11

Supplementary Table 8: **Evaluating different restraints and their effect on refinement results:** The HR restraints used in this paper (Step 4a) were validated by quantifying the strain of the cluster, electron density map flatness and comparison with the FBHR control. Additional checks were also performed with default restraints for Mn_4Ca cluster available in standard refinement packages. The results emphasize the suitability of the HR restraints (Step 4a) in modelling the Mn_4Ca cluster in the various S-states and indicate several issues with the default restraints.

	0F		1F		2F	
Restraint Type	Current Study; Custom	Default from CCP4 monomer lib (OEX)	Current Study; Custom	Default from CCP4 monomer lib (OEX)	Current Study; Custom	Default from CCP4 monomer lib (OEY)
Number of bond restraints	18	17	18	17	21	21
σ (Å)	0.05	0.02	0.05	0.02	0.05	0.02
Number of angle restraints	22	23	22	23	24	39
σ (degrees)	5.0	3.0	5.0	3.0	5.0	3.0
Restraints Z-statistics						
Bond RMSZ	0.94	2.78	0.83	2.33	0.43	1.59
|Z| > 2.0	0	10	0	7	0	5
Angle RMSZ	0.69	5.89	0.57	5.71	0.48	4.32
|Z| > 2.0	0	18	0	21	0	6
F_O-F_C map flatness (σ-scale)						
Signed Average	-0.8/0.9	-0.8/1.6	-0.4/0.8	-0.7/1.5	-0.5/0.8	-0.6/0.9
FBHR control	-0.8/0.9	-0.8/0.9	-0.3/0.7	-0.3/0.7	-0.5/0.9	-0.5/0.9
Flat solvent Control	-0.5/0.6	-0.4/0.6	-0.6/0.8	-0.5/0.7	-0.5/0.4	-0.5/0.4
Skew	0.21	1.11	0.19	0.94	0.16	-0.02
FBHR control	-0.05	-0.05	0.14	0.14	0.15	-0.15
Flat solvent control	0.21	0.29	-0.01	0.03	-0.26	-0.22
RMSD Statistics						
RMSD with FBHR (Å)	0.04	0.15	0.03	0.14	0.07	0.10

Supplementary Table 9: **B-factors of different oxygen ligands and water molecules close to the OEC in the S_3 state:** Note that all water molecules (except W1-W4) were modelled with a single component at 100% occupancy.

Oxygen/Water ID	**B-factor (Å^2)**
O1	22.8
O2	26.6
O3	29.5
O4	26.6
O5	28.3
O_X	30.5
W1	29.0
W2	27.3
W3	25.4
W4	29.7
W19	31.4
W23	24.9
W24	34.7
W25	26.7
W26	40.9

Supplementary Table 10: Mn-Mn distances from anomalous diffraction data and comparison with previous EXAFS measurements. The XFEL distances from monomer I in the 9.5 keV data are shown for the three di-μ-oxo bridges and the longer 3.3 Å bridge. The standard deviations for the XFEL data were calculated using the anomalous END/RAPID method (see Supplementary Methods). The EXAFS data is taken from Glockner et. al., JBC (2013)[9]. For the S_3 state, we used fit_A from that paper. For the ~2.7 Å distances, EXAFS fitting was done with either one distance (one shell) or two different distances (two shells, in the ratio of 1:1, or 1:2). In all three states S_1, S_2, and S_3, the two shell fits with ratios of 1:2 (total of three distances) were usually better, with two shorter distances ~2.7 Å and one slightly longer distance. We found that that for the S_3 state, the distances were typically more spread out (more distribution or heterogeneity) between the ~2.7 and ~2.8 Å distances. For the ~3.3 Å distances, the EXAFS fits are more difficult as both Mn-Mn and Mn-Ca distances contribute at this distance. We speculate that the larger discrepancy between the XRD and EXAFS results for the ~3.3 Å distances is because of this issue in the fitting procedure for EXAFS.

	XFEL crystallography	EXAFS
	R (σ (Å))	R
S_1		
Mn1 - Mn2	2.80 (0.02)	2.71, 2.71, 2.79
Mn2 - Mn3	2.82 (0.03)	
Mn3 - Mn4	2.78 (0.03)	
Mn1 - Mn3	3.22 (0.02)	3.28
S_2		
Mn1 - Mn2	2.79 (0.02)	2.74, 2.74, 2.74
Mn2 - Mn3	2.84 (0.03)	
Mn3 - Mn4	2.72 (0.03)	
Mn1 - Mn3	3.25 (0.04)	3.30

S_3		
Mn1 - Mn2	2.80 (0.02)	2.75, 2.75, 2.79
Mn2 - Mn3	2.83 (0.03)	
Mn3 - Mn4	2.76 (0.03)	
Mn1 - Mn3	3.41 (0.03)	3.26

Supplementary Table 11: Protein Data Bank ID and the corresponding DOI for the restraints to model the datasets shown in the paper. The raw X-ray free-electron laser diffraction images have been deposited in the Coherent X-Ray Imaging Database, www.cxidb.org (ID 246) with a DOI: 10.11577/3366809.

Dataset	PDB ID	Restraints DOI
0F (anomalous, 9.5 keV)	9Z33	10.5281/zenodo.17527367
1F (anomalous, 9.5 keV)	9Z34	10.5281/zenodo.17527403
2F (anomalous, 9.5 keV)	9Z35	10.5281/zenodo.17527428
0F (anomalous, 7 keV)	9Z36	10.5281/zenodo.17527432
2F (anomalous, 7 keV)	9Z37	10.5281/zenodo.17527442
0F (conventional, 9.5 keV)	9Z68	10.5281/zenodo.17604945
1F (conventional, 9.5 keV)	9Z69	10.5281/zenodo.17604967
2F (conventional, 9.5 keV)	9Z6A	10.5281/zenodo.17604979

Supplementary Notes

1. Isomorphous Difference Map Features: Figure 3a shows the isomorphous difference map peaks (contoured at 3.5σ) near the OEC for the 1F – 0F and 2F – 1F datasets. These maps can identify differences between 2 states with minimal model bias. Here we discuss these difference features in more detail and connect them to the structural changes observed. For the 1F – 0F data, feature A corresponds to a positive difference feature on Mn4 and highlights rearrangements in the Mn4 coordination environment. This includes a contraction of the Mn4-E333 coordination distance (shown in figure 4) as well as a slight tightening of the Mn4-W1/Mn4-W2 coordination distance. Feature B is due to movement of the Ca towards D1:E189 that leads to shortening of that interaction in the refined 1F model. In the 2F – 1F data, prominent positive/negative difference features are again observed on Mn4 (C, D) due to its movement in the expansion of the cluster in the 2F data. The signal for the binding of the new water, O_X, is seen in feature E as a 6.5σ difference peak. Features F and G are connected to the movement of Mn1 along the Mn1-Mn3 vector (also visible in the anomalous data) due to expansion of the cluster. Feature F also contains additional positive density due to motion of D1:E189 away from the OEC that creates room for insertion of O_X.

2. Monomer naming convention: Photosystem II is a homo-dimer, consisting of 2 monomers related by C2 symmetry. These are referred to as monomer I and II. Here we clarify the naming convention used in our work. In the current crystal form we define monomer I to be where the PsbV subunit makes a crystal contact with the PsbU subunit from a symmetry mate of monomer II. In monomer II, no such symmetry contacts are present in the lattice for its PsbV subunit. See Supplementary Figure 9 for visualization. This and previous studies from our group [1,2,4] have used uppercase chains to annotate monomer I and lowercase chains to annotate monomer II in the deposited data to the Protein Data Bank (PDB).

3. Map Terminologies: We list below the map terminologies and notations used in this paper for the convenience of the reader and to clarify the notations.

Anomalous Diffraction map: Also known as anomalous difference map or Δ_{ano} maps, they can be more formally defined as [$F_{obs}(+)$ - $F_{obs}(-)$] @ ($\varphi_0 - \frac{\pi}{2}$) where the phases are taken from the model itself (φ_0). $F_{obs}(+)$ and $F_{obs}(+)$ are the Friedel pairs.

Isomorphous Difference map: Also known as F_O-F_O map, these maps are calculated using the difference in structure factor amplitudes between 2 datasets and are phased with a common phasing model. In this paper, we use the notation B - A maps to denote isomorphous difference maps between 2 datasets A and B with the phases taken from dataset A. The B - A maps can be more formally defined as $[F_{obs}(B) - F_{obs}(A)] @ \varphi_A$

$2F_O – F_C$ map: These maps are also known as $2mF_O$-DF_C maps in the literature and are considered to be the best estimate of the electron density with minimal model bias and minimal residual error achievable through the maximum likelihood estimation of the structure factor amplitudes. Formally, they are denoted as $[2mF_{obs} - DF_{calc}] @ \varphi_0$ with the phases taken from the model (φ_0). The terms m and D are obtained from the maximum likelihood estimation of the structure during refinement.

$F_O – F_C$ omit map: These maps are also known as mF_O-DF_C omit maps in the literature and show difference features between the model and the experimental data. In this paper we calculated omit maps that showed the difference density by omitting key parts of the Mn_4Ca cluster from the model (Figure 3B). Formally they are denoted as $[mF_{obs} - DF_{calc}] @ \varphi_0$ where the phases (φ_0) and F_{calc} are calculated from a model after omitting the atoms of interest. Such maps are useful to determine light atom positions that are close to heavy metal atoms such as Mn.

4. Evaluating single-conformer 2F models: The final 2F structure consists of a mixture of S_2 and S_3 states, as determined by various spectroscopic methods (MIMS, EPR, Mn Kβ XES), and also based on crystallographic occupancy analysis of O_X shown in Figure 3. As an additional control, we explored how the simplest single-conformer model (with no O_X) would fit the observed electron density. The initial single-conformer refinement with HR restraints results in a 6σ F_O-F_C peak in the vicinity of Mn1/Ca (where we eventually model O_X). Multiple independent structures were then manually created with O5 moved towards the F_O-F_C blob (Mn1-O5 distances ranging from 1.75-2.50 Å, Mn4-O5 distances ranging from 2.5-4.0 Å). Additional starting structures were created where the E189 and O5 were moved closer to the F_O-F_C peak. The B-factors as well as the position of all the OEC atoms including O5 were perturbed (RMS=0.1 Å) in the starting models in order to explore the optimization landscape. The structures were refined using the FBHR restraints with a slight modification – no Mn-O5 restraints were included. In total, 143 independent single-conformer structures were refined. The final results (see Supplementary Figure 10) showed

the majority of the refined structures clustered around a conformation where the Mn4-O5 distance is ~ 2.2 Å, Mn1-O5 ~ 2.7 Å and in all those cases, the 6σ F_O-F_C peak remained. For refinement trials where the E189/O5 were simultaneously perturbed, the E189 conformation always moved back to the S_3 state conformation, as quantified by the dihedral angle and distance from the Ca. We also looked at structures that were not part of the majority cluster such as (a) where the Mn1-O5 distance was 2.2 Å (b) the F_O-F_C peak was the lowest (~3σ) and (c) The Mn4-O5 and Mn1-O5 distances were close to 2.4 Å. In all 3 cases, shown in Supplementary Figure 10, the refinements had diverged as evident from F_O-F_C peaks near the Mn sites as well as Mn-Mn distances that are incompatible with the anomalous data. Thus, all the structures that had converged in refinement had a 6σ F_O-F_C peak near Mn1/Ca. From this analysis, we conclude that a single-conformer model is not sufficient to explain the electron density map in the 2F state. The results support the two-conformer model in the 2F state with the O_X ligand bridging Mn1/Ca.

5. Effect of refining S_2 component in the 2F model: The multi-component 2F model is constructed by splitting the enzyme model into a S_3 component (65%) and a S_2 component (35%) in the active site region as discussed in the Methods section. The coordinates of the S_2 component were taken from the 1F model and kept fixed during structural refinement with only a group B-factor refinement of this component. This strategy is intended to reduce the refinement parameter space by fixing the known S_2 state, thus focusing the search space on the unknown S_3 state. However, we also checked the effect of the opposing strategy of refining the S_2 component simultaneously with the S_3 component on the final 2F model and O_X and OEC features. Two cases were investigated - (a) Refining the coordinates (xyz) of the S_2 component as is done normally for the 1F(S_2) dataset (b) Refining coordinates with limited freedom of the S_2 component by restraining them (including the protein part) to the known S_2 state structure using a tolerance level σ = 0.05 Å. This was done using the parameters within the *pdb_interpretation. reference_coordinate_restraints* scope in *phenix.refine*. In both cases, for the S_2 component, we used the same OEC restraints and anomalous Mn-Mn distances used in the 1F(S_2) refinement. The results near the Mn_4Ca cluster are summarized in Supplementary Figure 16 with representative changes/features. For (a) we did not observe any significant changes in the O_X electron density feature (omit map peak height ~8.5 σ) or geometry compared to the fixed S_2 case. The S_3 component overall did not change substantially although differences in Ca-Mn1 interactions were

observed (~0.05 Å). However, drastic changes were seen in the S_2 component where certain coordination distances and OEC geometry deviated significantly from the 1F(S_2) state. For example, the Ca-E189 distance changed from 2.79 Å to 2.98 Å. Another large change was observed for Mn3-E354 (1.93 Å to 1.75 Å. The Ca-Mn1 distance also changed (3.43 Å to 3.30 Å). These results highlight potential issues in simultaneously refining multiple components with little to no additional information to guide the optimization search space. In case (b) we kept the S_2 coordinates restrained to the known 1F(S_2) structure, thus providing additional information to guide the refinement. The effect of this is seen in the geometry of both components. In the S_3 component, the O_X electron density feature is still similar to the fixed S_2 case. The bond distances coordinating O_X are also unaffected. Changes in S_3 cluster geometry such as Ca-Mn distances are minor (~0.02 Å). For the S_2 component, the OEC geometry and coordination geometry are also now similar to the 1F(S_2) model. The changes in Mn and Ca coordination geometry noted in (a) are not observed here. From these tests, we conclude that it is necessary to either constrain or restrain tightly the coordinates of the S_2 component to obtain the best quality 2F model. This minimizes errors introduced into a known structure (S_2 component) that would in turn cause compensatory refinement errors in an unknown structure (S_3 component). We note that the strategy to keep the S_2 component fixed is also justified by the similar resolution of the 1F and 2F datasets.

6. Estimating occupancy of O_X and its error: The 2F model in this work is a 2-component model, similar to models published previously from our group, with the S_3/O_X population being 65%. The inefficiencies in S-state transitions, especially from $S_2 \rightarrow S_3$ are well documented, necessitating the splitting of the active site residues into 2 conformations. The S_3 population is similar to our previous estimates where normalized omit map peak heights of O_X along with in-situ Mn $K\beta_{13}$ X-ray emission spectroscopy (XES) measurements gave comparable estimates for O_X population. These estimates are also in line with ex-situ EPR and O_2 evolution measurements. The omit map peak heights shown in Supplementary Figure 8 are in agreement with the previous publications from our group and are consistent with a ~65% O_X occupancy. Here we provide additional occupancy analysis of the crystallographic data. We also discuss the error estimate of O_X occupancy and signal.

A. Refinement of O_X occupancy in a 1-component model: We refined the O_X occupancy using a 1-component model in *phenix.refine*. 100 independent refinement trials were performed to

disentangle the correlation between B-factor and occupancy during refinement. Prior to refinement, the O_X occupancy was randomized between 0 - 1 and the B-factors of all the protein/OEC atoms were shaken. The refinement was done using HR restraints, similar to what was used for the final 2F model. All other refinement settings were also the same as those used for the 2F model. The occupancies of all other OEC atoms were kept at 1.0. The final O_X occupancy estimate is 0.65 ± 0.08, based on 57 structures where the refinement converged (occupancy < 1.0). For the remaining 43 structures, the O_X occupancy refined to exactly 1.0, with abnormally low B-factors and negative F_O-F_C density at its position, indicating the occupancy needs to be less than 1.0. The Mn1-O_X distances in these refinement trials were also unphysically short (< 1.4 Å), indicators of a diverged refinement. This is not surprising as this type of refinement approach can be highly unstable because of the over-simplified nature of the 1-component model as well as the weak O_X signal compared to the other bridging oxygens and Mn.

B. Scanning a set of occupancy scenarios in 2-component models with O_X: In this approach, we constructed 2-component models (S_2 and S_3) where the population of S_3 (containing O_X) was varied from 0 - 100% in steps of 10%. The S_2/S_3 populations in these refinement trials were kept fixed and only coordinate/B-factor refinement was carried out. The hypothesis here is that the optimal ratio of S_2:S_3 population would yield an F_O-F_C density ~ 0 at the O_X position. These refinements were done with FBHR restraints in order to allow the model maximum freedom. As shown in Supplementary Figure 18, at low occupancy of S_3, a strong F_O-F_C peak is observed which gradually reduces with increasing S_3 population. At 60%, the F_O-F_C peak is ~ 0, suggesting that the O_X population is in this range.

C. Error of the O_X signal, occupancy and geometry using bootstrapping of lattices during merging: We used the bootstrapping approach (see details in 'Errors estimation on bond lengths and occupancy' in Supplementary Methods) to evaluate the uncertainty in the O_X signal, occupancy and geometry. To this end, 3 different refinement trials were run for each of the 100 resampled mtz datasets (so 300 in total).

(a) We ran 100 refinement trials with no O_X in a 1-component model to estimate the residual F_O-F_C signal near Mn1/Ca. We obtained a value of 6.0 ± 0.5 σ. This estimate is similar to the results shown in Supplementary Figure 10. This demonstrates the presence of a strong difference density feature in the Mn1/Ca coordination environment that is reproducible at a data merging level.

(b) Refinement trials with O_X in a 1-component model were performed and occupancy refinement enabled for O_X only. Here we obtained an O_X occupancy of 0.64 $\pm$ 0.07, very similar to the value obtained in section A above. This provides additional support to our O_X occupancy estimate.

(c) Two-component models were refined where the S_3/O_X occupancy was kept fixed (65%) in order to obtain error estimates on the bond lengths with the bootstrapping procedure. The standard deviations on the bonds are 0.08 Å (O5-O_X), 0.10 Å (Mn1-O_X) and 0.06 Å (Ca-O_X). These estimates are less than the END/RAPID estimates for the same bonds (0.11 Å , 0.10 Å and 0.09 Å). This trend is expected as END/RAPID errors are an upper limit of the true errors due to the nature of the perturbations introduced during calculations. The error calculations further reinforce the final O_X geometry (Figure 4), and provide complementary evidence that the O-O bond is not formed in the S_3 state.

Supplementary Methods

Theoretical X-ray scattering map values for 2- or 3-atom models: We compared $2F_O$-F_C map traces (Mn-O5-O_X, O_X-Mn1-O, Ca-O_X) for the 2F dataset with traces of the theoretical fourier images of the electron density from X-ray scattering for linear two or three-atom models (Mn-O-O or O-Mn-O or Ca-O). The idea behind the computation is to use a simplified tunable model to set expectations on how the error-free Fourier map profiles look like under different conditions at the resolution of the data and compare them with the experimental data (See Supplementary Figure 14 for example). For our case, we wanted to see how the map profile is affected by the presence or absence of O_X. The theoretical electron density maps were generated using the bond distance and B-factor parameters provided in Supplementary Figure 15 at 1.93 Å resolution. The Mn atoms (purple) and the oxygen atoms (in red) were modelled at 100% occupancy while O_X (in yellow) was varied in occupancy (0, 25, 50, 75, 100%). For the O5-O_X trace, a Mn atom was used as a proxy atom for the Mn3-O4-Mn4 bridge with the B-factor adjusted to match the main peak height. The *maptbx.map_curves* routine in the PHENIX package was used for the calculations within a custom python script (See Code availability section). The theoretical Fourier maps were converted from absolute values ($e^-/Å^3$) to σ scale by using the same conversion factor between the two scales in the experimental map. For the experimental map, the absolute scale was determined by using an estimate of F_{000} from the program *phenix.f000* (3056554.731 e^-). The data is shown in Figure

3c and discussed in the main text. The results indicate the presence of a new ligand O_X in the 2F data with an occupancy in the range of 50-75% best match the experimental profiles. We note that a precise population estimate is best done using occupancy refinement or omit map peak height comparison which have been shown in Supplementary Figure 8. For the final model building, we used an O_X (and S_3) population of 65% in the 2F modelling, consistent with the spectroscopic analysis. This is also similar to the population used in previous publications[1] within the limit of crystallographic detection (+/- 10%).

We also calculated F_{calc} maps of the 3-atom or 2-atom models from the refined structure itself using the same set of miller indices as the $2F_O$-F_C map in the same unit cell as the 2F dataset. This would allow us to further confirm the analysis and check for the effect of the environment and data completeness on these maps. These models were created by removing all but the relevant 2 or 3-atoms of interest from the 2F structure. We retained the 2 components in the model and only changed the occupancy of O_X between 0-100% (as done for the theoretical traces). For the O5-O_X trace, the Mn3-O4-Mn4 bridge was part of the model for F_{calc} calculation. For Mn1-O_X, the carboxylate oxygen of D342 bridging with Mn1 was part of the model. All B-factors and coordinates were kept the same as in the 2F structure. The tool *phenix.fmodel* was used with default settings to calculate the F_{calc} maps for different O_X occupancies. The F_{calc} maps were scaled the same way as described above to bring it to the same scale as the experimental map. The results, shown in Supplementary Figure 12 further support the modelling of the ligand O_X in the 2F dataset.

Validation of Restraints: In order to ascertain that the harmonic restraints and refinement methodology used in this work are not biasing the OEC model towards a conformer that is not supported by the experimental data, we ran multiple validation checks. We also wanted to establish the level of consistency between the results obtained using the current restraints (HR set) and those using our previously published restraints[2] as well as the default restraints available for ligand entries OEX and OEY in the refinement packages in the Phenix and CCP4 suites. Since it is currently not possible to deposit our custom OEC restraints and have them available on the PDB sites (both used in this paper as well as our previous papers), we have made them available in the Supplementary Data 1 and 2 as well as on Zenodo (see Data Availability section). We reiterate that our publications (current and previous) did not use the default OEC restraints (for OEX and OEY) that are available in the current refinement packages.

With anomalous data (Step 3, see Supplementary Figure 7), we ran refinement with HR (Step 3a, σ = 0.05 Å for bonds) and for validation, separate refinement with Flat bottom harmonic restraints (FBHR, step 3b, no penalty for Mn-O in the range of 1.6 – 2.3 Å). With conventional data (Step 4), we ran refinement with HR (Step 4a, σ = 0.05 Å for bonds) and for validation, refinement with FBHR (Step 4b, no penalty for Mn-O in the range of 1.6 – 2.3 Å). We also ran independent refinement with restraints used in our previous publications[1] (Step 4c, σ = 0.05 Å for bonds), default restraints from standard packages like CCP4 (Step 4d, σ = 0.02 Å for bonds) For steps 4c and 4d, we used the Mn-Mn restraints as defined in the restraints file instead of the anomalous data. In all cases, interactions of the Mn/Ca with the coordinating amino acids were defined using very loose harmonic restraints with σ = 0.50 Å. The following validation statistics are shown below to demonstrate the suitability of the HR restraints used in this paper -

A) *RMSZ and individual Z-scores for the OEC restraint terms after refinement*: Z-scores for bond restraints are defined as $\frac{d-d_0}{\sigma}$, where d is the observed bond distance, d_0 is the ideal distance and σ is the expected standard deviation on the distance. They are a normalized measure of the variation of the model from the expected values. Angle Z-scores can be defined in a similar manner. Individual absolute Z-score values > 2.0 would indicate the X-ray data is in disagreement with the restraint term and there is substantial penalty during refinement. RMSZ values ∼1.0 for the cluster would indicate the restraints overall are in good agreement with the experimental data and are appropriate for modelling.

B) *F_O-F_C map flatness*: If restraints push the atoms towards a conformation that disagrees with the X-ray data, this would also be reflected in the *F_O-F_C* difference map, either as positive or negative density above a certain noise level. We defined map flatness to be the average signed *F_O-F_C* electron density (positive and negative values kept separate) and the skew value within a box of 5x5x5 Å about the center of mass of the 4 Mn atoms. As a control, the same metrics were calculated from FBHR refinement (which would minimize the map flatness with the geometry term downweighed) and also bulk solvent regions where no structured density exists.

C) *RMSD of model distances between HR and FBHR refinement cases*: FBHR restraints have no penalty for Mn-O bond distances between 1.6 – 2.3 Å and beyond that range, σ = 0.05 Å value is used. Thus, the weight of the OEC restraints is downweighted considerably in FBHR refinement. This serves as a control to evaluate if the X-ray data prefers an OEC conformation that is very

different from the description in the HR restraints. The level of agreement is quantified by the RMSD of the bond distances between the two sets of refinement.

The results for the 0F, 1F and 2F cases from step 4 with HR restraints (Step 4a) are summarized in Supplementary Table 8. We also present additional calculations with the default restraints (Step 4d). The restraints report shows no evidence for the restraints from our work biasing the model towards a conformation not supported by the X-ray data. The results from the HR refinements are within the acceptable limits as shown in the control cases. We also clarify that no assumption of oxidation states for the Mn is necessary to do the crystallographic model (both in this paper and our previous papers). It is noteworthy that the default restraint set shows strong disagreement with the X-ray data during refinement.

Errors estimation on bond lengths and occupancy: Determining the effect of measurement errors on the final bond lengths is still an active area of research[3]. The error models used in serial crystallography for merged structure factor intensities either underestimate or overestimate the true measurement error with the problem being particularly severe in the highest resolution bins. Consequently, it becomes harder to precisely evaluate the effect of intensity measurement errors on the final structure and biological interpretation. Our group's practice has been to establish an upper bound of the structural error by using the END/RAPID procedure [1,2,4]. The method allows us to propagate the intensity measurement errors to errors in the structural model. In this paper, we used two different strategies for the error propagation with the END/RAPID procedure depending on which set of intensities were used (anomalous vs. conventional)[5]. In addition, we used a third error estimation method with a bootstrapping approach to determine uncertainty in O_X signal and occupancy[6].

Conventional Intensities (END/RAPID): The steps involved are as follows – We first perturb the scaled structure factors (F_{obs} scaled to F_{calc}) by a random amount in the range +/- $|F_{obs} - F_{calc}|$. This range is an upper bound on the possible measurement error. We create 100 synthetic datasets in this manner and run 100 independent refinements. The starting structural coordinates in each refinement trial are also perturbed by a random amount using the modify_start_model phil parameter in *phenix.refine*. The Mn-O harmonic restraints for the OEC during the END/RAPID step are relaxed (σ = 0.50 Å) along with the Mn-Mn restraints from the anomalous data (σ = 0.05 Å).

Supplementary Table 6 shows the comparison of error estimates for select OEC bond lengths from the current 0F data (1.97 Å) with a previous 0F dataset (2.08 Å, PNAS 2020). Both the Mn-Mn and Mn-O bond length errors are lower in the current data, highlighting the effect of improvements in both the resolution and the modelling procedure. On average, the Mn-O error from the current 0F dataset is ~ 0.08 Å and it is 0.13 Å for the previous 0F dataset. Supplementary Table 7 provides the error estimates on key bond lengths in the three flash states (9.5 keV) discussed in the paper.

Anomalous Intensities (END/RAPID): The Anomalous difference signal is the difference between the Friedel mates ($I_{obs}(+)$ and $I_{obs}(-)$) and is not the difference between an observation and a model. Thus, larger F_{obs}-F_{calc} perturbations for error estimation are not appropriate for estimating errors related to Mn position from anomalous data. For this part, we used $\sigma(I_{obs}(+))$ and $\sigma(I_{obs}(-))$ as standard deviations for perturbing the $I_{obs}(+)$ and $I_{obs}(-)$ intensities respectively and generating 100 synthetic datasets. The rest of the procedure is similar to what is detailed above for conventional intensities. These results are shown in the Mn-Mn errors in Figure 2 and Supplementary Figure 2.

Conventional Intensities (Bootstrapping): We evaluated the robustness of the O_X signal in the 2F dataset by creating 100 resampled datasets by randomly drawing lattices with replacement from the pool of available lattices until the same number of lattices as in the original dataset was reached[6]. In the context of serial crystallography, the term "lattice" refers to the collection of unmerged raw intensity measurements of Bragg spots arising from a single crystal. The resolution was kept to be the same throughout (1.93 Å for the 2F dataset). This algorithm was implemented in the *cctbx.xfel.merge* program and can be accessed via the phil option *filter.algorithm=bootstrap*. The 100 datasets were created by setting the random seed to range from 1 to 100 (phil option *filter.bootstrap.random_seed*). This technique tests the reproducibility of the O_X signal at the data merging level, thus providing a complementary check to our END/RAPID method (which resamples structure factors after the merging step). Once the 100 datasets are created, structural refinement was carried out using *phenix.refine*. We ran three sets of refinement trials (300 in total) - (a) without O_X, to estimate the error in F_O-F_C signal near Mn1/Ca using a 1-component model (b) with O_X occupancy refined to estimate the uncertainty in O_X population using a 1-component model (c) with O_X occupancy fixed (65%), as part of a 2-component model, to estimate the error in the O_X geometry. The results are described in Supplementary Note 6. The error estimates on the

O_X bond lengths and Mn-O bridges are slightly smaller compared to those obtained from END/RAPID (Conventional Intensities, END/RAPID).

Flat-bottom Harmonic Restraints (FBHR) Implementation: The FBHR restraints were used in the program *phenix.refine* by defining the restraints in an edits file (file format specific to *phenix.refine,* see example Supplementary Data 4). For bond pairs, one can specify atom 1 and atom 2 using the PHENIX selection syntax, the expected distance (parameter *distance_ideal*) and the standard deviation (parameter sigma). The range over which no penalty is applied i.e. the flat-bottom part is specified using the parameter *slack* which defines the flat-bottom part to be *distance_ideal* +/- *slack*. The penalty from the standard deviation, *sigma*, is applied on either side of the flat-bottom part. The FBHR restraints term can be compared to the standard Harmonic restraints (HR) term in the residual plot shown in Supplementary Figure 13.

Map Calculations: $2F_O - F_C$, $F_O - F_C$ and anomalous diffraction maps shown in this paper were calculated from map coefficients written out by *phenix.refine* [7]. For 1D traces shown in this paper (Figure 2c, 3c), custom Python scripts using the phenix software were developed. These have been made available at https://github.com/asmit3/eden with instructions therein to run these scripts. The 1D traces were obtained using a grid spacing of approximately 0.2 Å for all the datasets (resolution_factor=0.1 in *CCTBX*). Omit maps shown in Figure 3b were calculated by the program *phenix.polder* and subsequently displayed using the normal omit map coefficients. Isomorphous difference maps were calculated using the *phenix.fobs_minus_fobs_map* program with default settings. Map centroids shown in Supplementary Table 5 were calculated around a 1.0 Å sphere of the final Mn positions using maps that were generated prior to addition of Mn (i.e., step 1). Custom scripts to obtain map centroids are also on the GitHub page. For the Polder $F_O - F_C$ omit maps[8] (Polder omit map coefficients) shown in Supplementary Figure 17, a few additional steps were implemented to mitigate any phase bias from the OEC and the surroundings. They were as follows - (1) Delete the OEC and all first shell amino acids/waters from the 2F model in coot. The amino acids included chain A/a residues 170, 189, 332, 333, 342, 344, 61, 337, 190, 161, 165, 185 and chain C/c residues 354, 357 (2) Randomize the xyz and B-factors of the resulting model (3) Refine for 100 cycles to lose memory of previous phases (4) Insert the deleted residues into the new model file as a placeholder for the Polder calculation (5) Calculate the Polder omit map using *phenix.polder* by omitting the entire region that had been deleted.

Supplementary References

1. Kern, J. *et al.* Structures of the intermediates of Kok's photosynthetic water oxidation clock. *Nature* **563**, 421–425 (2018).
2. Bhowmick, A. *et al.* Structural evidence for intermediates during O_2 formation in photosystem II. *Nature* **617**, 629–636 (2023).
3. Mittan-Moreau, D. W. *et al.* Robust error calibration for serial crystallography. *Acta Crystallogr D Struct Biol* **81**, 265–275 (2025).
4. Ibrahim, M. *et al.* Untangling the sequence of events during the $S_2 \rightarrow S_3$ transition in photosystem II and implications for the water oxidation mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 12624–12635 (2020).
5. Lang, P. T., Holton, J. M., Fraser, J. S. & Alber, T. Protein structural ensembles are revealed by redefining X-ray electron density noise. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 237–42 (2014).
6. Grünbein, M. L. *et al.* Effect of X-ray free-electron laser-induced shockwaves on haemoglobin microcrystals delivered in a liquid jet. *Nature Communications* **12**, 1672 (2021).
7. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D* **75**, 861–877 (2019).
8. Liebschner, D. *et al.* Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Crystallogr D Struct Biol* **73**, 148–157 (2017).
9. Glöckner, C. *et al.* Structural Changes of the Oxygen-evolving Complex in Photosystem II during the Catalytic Cycle. *Journal of Biological Chemistry* **288**, 22607–22620 (2013).