Online supporting materials (three supporting sections, one supporting table, three additional supporting references, and eight supporting figures) for

Heterogeneous composition of oxygen-evolving complexes in crystal structures of dark-adapted photosystem II

Jimin Wang[1,*], Christopher J. Gisriel[2], Krystle Reiss[2], Hao-Li Huang[1], William H. Armstrong[3,4], Gary W. Brudvig[1,2], Victor S. Batista[2]

[1]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520-8114. [2]Department of Chemistry, Yale University, New Haven, CT 06520-8107. [3]Department of Chemistry, Boston College, Chestnut Hill, MA 02467. [4](Footnote: Retired).

1. Deviation of spherical symmetry and accuracy of analytically fitted parameters of metal centers in nitrogenase structures

The FeMo (M) and P clusters within FeMo nitrogenase bear similar features to the OEC in PSII. Fe is one atomic Z number greater than Mn, 26 versus 25. Fe centers in the P-cluster of nitrogenase have a formal oxidation number of II, thus containing 24 electrons in theory. However, the Fe5 and Fe6 centers exhibit two different conformations (Fig. S2); therefore, we neglect these atoms from the 8 Fe atoms found per P-cluster for this analysis.[15] Fe centers in M-clusters are divided into two groups with their formal oxidation numbers different by one or less, a situation more similar to the Mn centers in the OECs of dark-adapted PSII dimers.

Inspection of omit F_o-F_c maps of the 3u7q structure shows that the electron density (ED) peaks are spherically symmetric with a visual dynamic range from 5σ to 80σ (Fig. S2).[16] Visually, these peaks are spherically symmetric (Fig. S2). At a contouring level below 5σ, noise peaks appear. Quantitative analysis of omit F_o-F_c ED peaks in two steps is demonstrated in Figure S2. The ED was then spherically averaged and the standard deviation relative to the average value was plotted (Fig. S2E). This deviation is due to anisotropic atomic motions and intrinsic asphericity of the atomic ED distribution. It is relatively small and very similar among the 6 Fe centers of the P-clusters analyzed. Analytic fitting parameters go through an internal normalization step of the ED value, $\rho(r)/\rho(0)$, which removes its dependence on either occupancy or relative EN (Fig. S4C). The variance $<|\Delta r|^2>$ of the ED distribution or atomic B-factor ($B=8\pi^2<|\Delta r|^2>$) depends only on the half-width of the ED peak. Any small difference of the variance between the query and reference can be removed by least-squares fitting of the slope in a logarithm plot (Fig. 4D). After a small ΔB is fitted, the relative EN can be analytically fitted in a second step. Using these two parameters, all curves become nearly completely superimposed (Fig. S4B), implying that the ENs determined are accurate.

Two types of standard deviation were calculated for 6 Fe centers (Fe1, Fe2, Fe3, Fe4, Fe7, and Fe8) of the P-cluster in the 3u7q structure at 1.0-Å resolution: (i) the 3u7q model without any re-refinement (σ_1 in Table S1), and (ii) the re-refined 3u7q model (σ_2 in Table S1).[16] In addition, pairwise differences were calculated between the two 3u7q models before and after model re-refinement (σ_3 in Table S1). An analysis of the same kind has been carried out for the 4wes structure at 1.08 Å resolution.[15] These analyses have shown that relative ENs determined using this method are within 2% and not very sensitive to the quality of atomic models used for peak decomposition for removal of contribution of ligands to the ED peak, nor to ripple effects of Fourier series termination. In the main text, we only presented the results of analysis for the Fe1, Fe2, Fe3, and Fe4 centers although the analysis was done for all 8 Fe centers (Fig. 1). With the exception of the Fe5 and Fe6 centers, the variations for the remaining Fe7 and Fe8 are also very small.

2. Deviation of spherical symmetry and accuracy of analytically fitted parameters in PSII structures

In two 6jlj-based simulations described in the main text,[18] X-ray atomic scattering factors used for the 8 Mn centers of the two OECs are identical as is the distribution of the 8 resulting ED peaks in omit F_o-F_c maps. In this situation, when ED peaks of the query and reference are aligned using two fitted parameters obtained within a small sphere of 0.7 Å radius, the distribution of these two ED peaks is also perfectly aligned in larger spheres everywhere in real space as long as their ED values are measurable above random errors. A difficulty in the experimental study is, however, that atomic scattering factors of Mn in an unknown oxidation state are not known, nor are they in mixed oxidation states. They may differ from an internal reference. In this case, an alignment of the straight-line region of logarithm density plot against distance squares within a small sphere between the query and reference does not guarantee that the remaining regions in large spheres will also be aligned. The tails of these plots may curve differently downwards, which could affect the relative ENs determined. To determine the extent of errors of this kind, we carried out two analyses in parallel, one within a sphere of 0.7 Å radius and the other within 1.2 Å for Mn ions of the OEC of the 6jlj structure[18] We found that the relative ENs determined were reproduced within 0.001 units even though the absolute ENs determined between two spheres are different for the Mn2 reference.

Another concern arises as to how very large model R-factors for the five structures analyzed in this study might affect the resulting relative ENs. Large model R-factors are often associated with modeling errors elsewhere in the structure and away from the OEC. We carried out an error-removal procedure by zeroing all grid points (or voxels) outside a sphere of 4-Å radius of the Mn center of interest in the 6jlj structure, inverted the ED peak (with surrounding negative ripple effects of Fourier series termination retained), and recalculated the ED peaks for analysis. A comparison of results with and without this error-removal procedure shows that the relative ENs obtained were again reproduced to within 0.001 units. This observation suggests that, because modeling errors of this kind were distributed away from the OECs in omit F_o-F_c maps, they affected both the query and reference the same way so that they were completely cancelled out.

Another issue is a deviation of the ED distribution from spherical symmetry (or asphericity) associated with anisotropic motions (Fig. S5). We examined ±1 standard deviation to the spherically averaged mean value of the ED distribution for the metal ions of OEC in the 6jlj structure (Fig. S5A-S5C).[18] It is evident that deviation from spherical symmetry in the OEC of PSII is much larger than that in the P-cluster of the nitrogenase structures examined (Fig. S3, S4). A larger deviation is likely associated with anisotropic motions of the entire cluster according to preliminary analysis of anisotropic tensors present in the coordinate file. In fact, these tensors were derived from translation-liberation-screw (TLS) motions, not actually individually refined B-factor tensors. After fitting of two parameters, it was observed that the standard deviations from spherical symmetry for the four Mn centers are nearly identical, and that of Ca center was slightly larger (Fig. S5A-S5C).

In addition to the classic omit F_o-F_c amplitude difference maps for peak decomposition, omit F_o-F_c vector difference maps were also calculated for analysis in which X-ray form factors of neutral Mn atoms were used for calculation of the F_o map. In this approach, the total EN for each Mn center is assumed to be the same 25 electrons, which has an anti-model bias issue, i.e., if the differences in ENs do exist between the query and reference, the ENs obtained will appear smaller than both their true ENs and those obtained using the omit F_o-F_c maps. The relative ENs were determined using analytic fitting procedures with and without an additional flattening procedure in parallel, and it shows that the relative ENs for Mn centers were reproduced between the two procedures to within 0.004 units and Ca centers to within 0.010 units. However, a comparison between the omit F_o-F_c *amplitude* difference maps and the omit F_o-F_c *vector* difference maps reveals that the relative ENs determined for Mn centers were reproduced to within 0.010 units for the Mn1 center and about 0.04 units for the Mn3 and Mn4 centers (up to 0.8 electrons), and the Ca center to within 0.080 units (1.7 electrons). These represent the largest errors in all of the tests for computational reproducibility of the relative ENs.

For the Ca center, differences in atomic scattering factors of different Gaussian terms (see main text) between Ca and the reference Mn2 center may likely play a role in such a large irreproducibility. For the Mn3 and Mn4 centers, there might exist a mixture of oxidation states that could not be fully modeled using Gaussian motions of single Mn centers even though the identification of multiple centers remains not possible from F_o-F_c maps at this resolution.

In the omit vector difference maps after expanding to P1, deviation from spherical symmetry in the ED peak distribution in real space reflects a deviation from spherical symmetry in the amplitudes of the structure factors in reciprocal space (Fig. S5D-S5F). According to the standard deviation of structure factors from their spherically averaged mean values, an estimated error on the absolute EN determination is 0.42 *e* for Ca^{2+} ion (or 2.2%) and 0.60 *e* for each of four Mn ions (2.3%) (Fig. S5D-S5F). However, an anisotropy of this kind has again been largely cancelled out in our relative EN determination.

3. Potential experimental sources for errors of relative electron numbers determined analytically

There are four main types of systematic errors that can affect the accuracy of the absolute ENs determined: (i) incomplete models due to missing ordered solvent or lipid molecules, i.e., poor overall quality of models,[73] (ii) ripple effects of Fourier series termination due to missing measurable high-resolution amplitudes,[59] (iii) anisotropic motions of the OEC as a single group because, at about 2.0-Å resolution, refinement of individual anisotropic tensors is difficult and inaccurate so that anisotropy of atomic motion could not be fully corrected for, and (iv) different rates of XFEL-induced destruction.[40,74,75] Missing ordered water and lipid molecules distributed far away from the OEC could affect the ED distribution of omitted metal ions of the OEC largely through a resolution-dependent figure-of-merit weighting manner, which would be completely cancelled in a relative EN determination. If metal ions of similar kind have similar or the same B-factor, the effects of missing unmeasurable terms have ripple effects of the same magnitude, and again, their effects would largely be cancelled out. Similarly, given anisotropy results from TLS motions, all anisotropic tensors of metal ions within each OEC are perfectly aligned. As a consequence, deviations from spherical symmetry in omitted ED peaks are almost identical for all five metal ions and can be visualized by the standard deviation of spherical

averaging to the mean value (Fig. S3A-S3C). On a relative scale, the contribution of anomalous electrons can be completely ignored when data are collected away from *K*-edge of metal ions.

In the omit F_o-F_c maps examined in this study, the contribution of neighboring atoms has been removed using the known atomic models, but the ENs obtained for omit Mn centers and the heights of the resulting ED peaks are only about half of their true values[30] so that our methods are not suitable for determination of the absolute ENs. In the previous studies using σ_A-weighted $2F_o$-F_c maps,[12-14] the contribution of neighboring atoms was not directly removed because it was very small. As a consequence, the resolution required was much higher than in the approach used in this study, and the underlying atomic B-factors should be much smaller so that there was sufficiently intrinsic spatial resolution of neighboring ED peaks with no need of peak decomposition.

Table S1. Standard deviation of relative ENs as a function of added B-factor with decreasing effective resolution estimated.[a]

ΔB (Å^2) added	0	10	20	30	40
Effective Resolution (Å)	0.78	1.43	1.90	2.28	2.62
σ_1(relative ENs) (%)[b]	1.07	1.13	1.29	1.43	1.52
σ_2(relative ENs) (%)[c]	1.11	1.24	1.41	1.50	1.53
σ_3(relative ENs) (%)[d]	0.99	1.33	1.62	1.78	1.87

Footnote:

a Effective resolution was estimated with the Wilson B-factor, which was an empirical relationship upon analysis of all X-ray diffraction data sets deposited in the PDB (reference 75). Using mean atomic B-factor for Fe ions (~ 5 Å^2). Signal-to-noise ratios for Fe ions in the highest resolution shell of 1.0 Å in the 3u7q data set were much higher than those of protein atoms with an estimated effective resolution much higher than the resolution limit of the diffraction data.

b Standard deviation of relative ENs relative to the mean value was calculated assuming that all 6 Fe centers have the same EN values. The summed occupancy of D7499/Fe1 was 1.80 between two conformations in the 3u7q coordinate file (0.80 + 1.00) (which is an error) while it was 1.0 for the remaining Fe centers (0.80 + 0.20). This summed occupancy was fixed to 1.00 (thus this important error was corrected)during calculation of omit F_o-F_c maps for other Fe centers. The mean free R-factor was 19.44 ± 0.04% for 12 Fe-omit models without any refinement.

c Standard deviation of the ENs relative to the mean value was calculated after 5 runs of re-refinement. The mean free R-factor was reduced to 15.70 ± 0.03% for 12 Fe-omit models with an additional 30 steps of refinement iteration. A slightly increased standard deviation with improved models implies that differences between different Fe centers are not random noise.

d Standard deviation of variations of relative ENs between two sets of models with and without re-refinement.

Additional Supporting References

73. Wang J (2017) Systematic analysis of residual density suggests that a major limitation in well-refined X-ray structures of proteins in the omission of ordered solvent. Protein Sci. 26:1012-1023.

74. Lunin VY, Grum-Grzhimailo AN, Gryzlova EV, Sinitsyn DO, Petrova TE, Lumina NL, Balabaev NK, Tereshkina KB, Stepanov AS, Krupyanskii YF. (2015) Efficient calculation of diffracted intensities in the case of nonstationary scattering by biological macromolecules under XFEL pulses. Acta Cryst D71:293-303.

75. Wang J (2017) Experimental charge density from electron microscopic maps. Protein Sci 25:1619-1626.

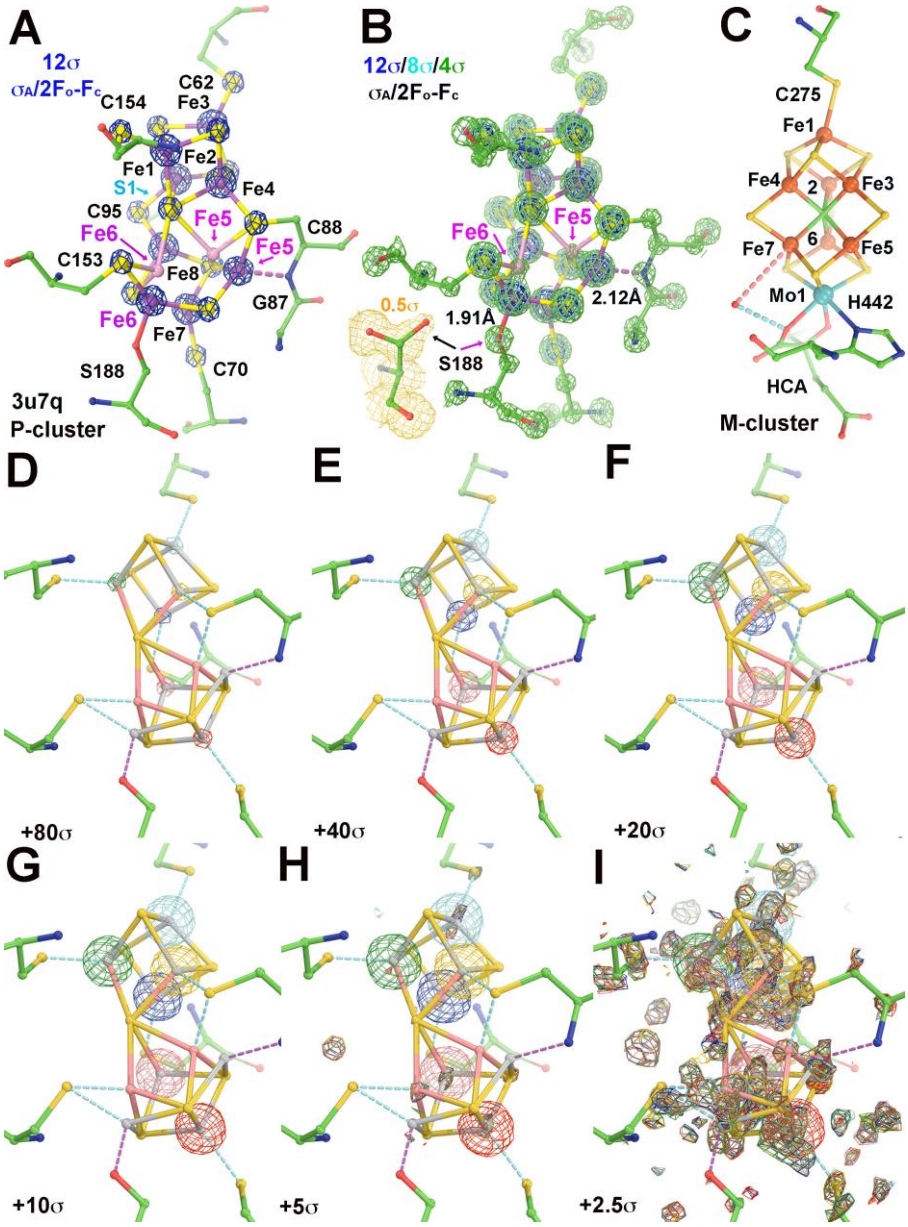

Figure S1. P-clusters and M-clusters of the 3u7q nitrogenase crystal structure and omit F_o-F_c difference Fourier peaks. (A) P-cluster superimposed with σ_A-weighted 2F_o-F_c ED map contoured at 12σ. Both Fe5 and Fe6 centers (in magenta) have two alternative conformations. (B) At reducing contour levels of 4σ (forest green), minor conformations of both Fe5 and Fe6 centers are visible although not at 8σ (cyan) and 12σ (blue). Minor conformations of ligands (such as S188 sidechain) were also clearly detectable although some of them have not been modeled. (C) M-cluster with 7 Fe centers plus one Mo center and homocitric acid (HCA) as one ligand to Mo. (D-I) Omit F_o-F_c maps contoured at 80σ (D), 40σ (E), 20σ (F), 10σ (G), 5σ (H), and 2.5σ (I).

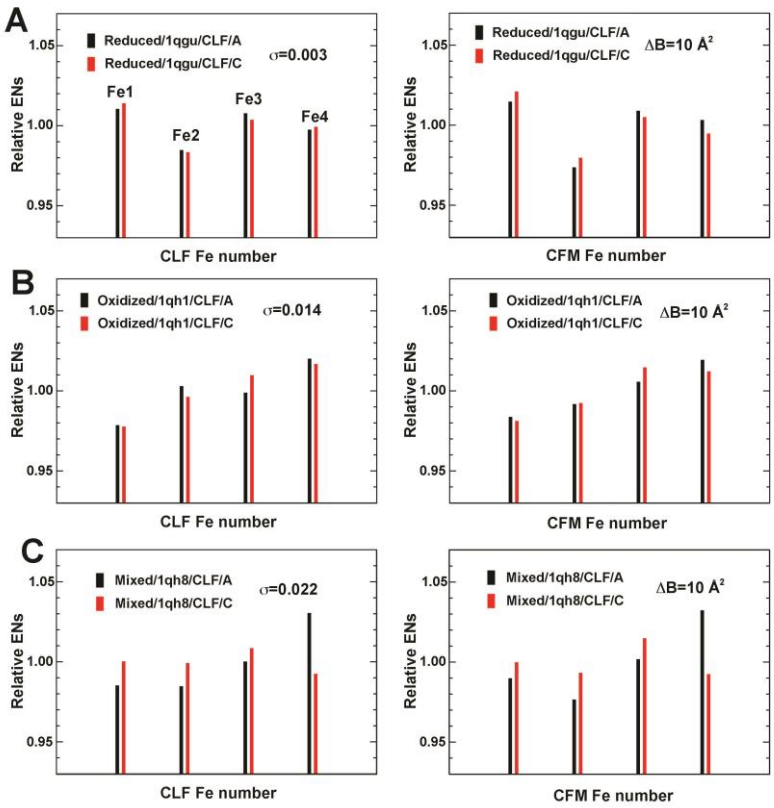

Figure S2. Comparison of relative ENs determined for Fe centers between two independent P-clusters in three nitrogenase structures at 1.60-Å resolution. (A) The chemically reduced state of the 1qgu structure. (B) The chemically oxidized state of the 1qh1 structure. (C) Native mixed oxidation state of the 1qh8 structure. Left column: results of analysis on the original omit F_o-F_c maps. Right left: $\Delta B = 10$ Å^2 was added in the omit F_o-F_c maps.

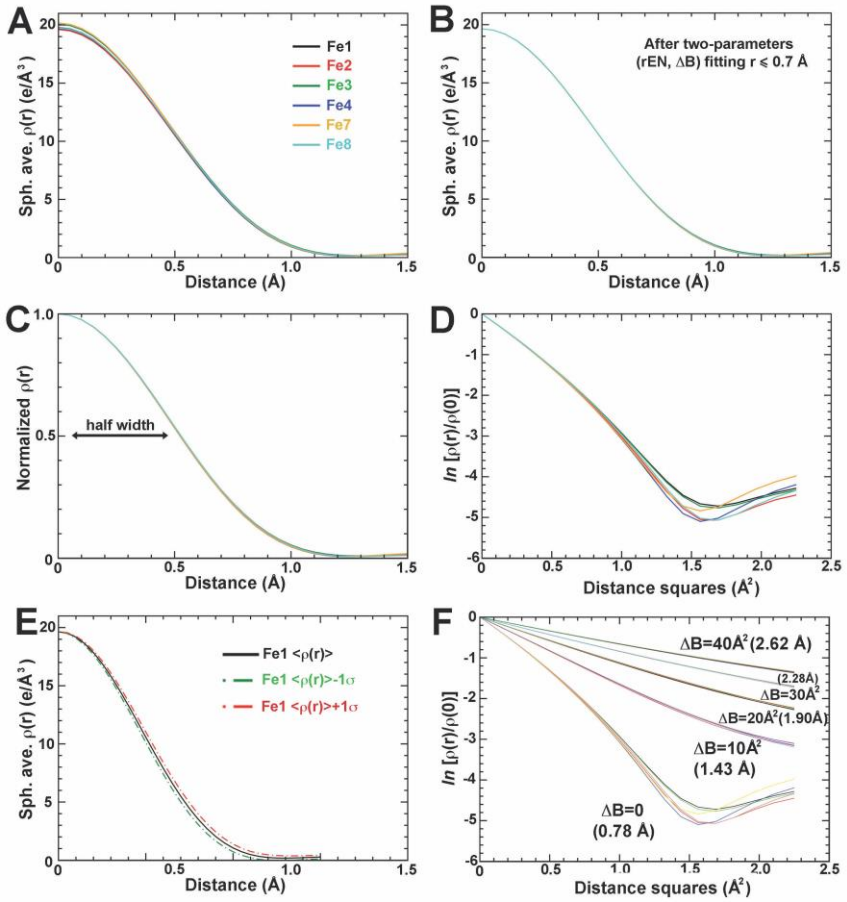

Figure S3. Quantitative analysis of omit F_o-F_c ED peaks for 6 Fe centers of one P-cluster of the 3u7q nitrogenase structure. (A) spherically averaged $\rho(r)$ as a function of r. (B) Results after two-parameter (relative EN and different B-factor) fitting within a radius of 0.7 Å. (C) The first step of fitting is through an internally normalized ED $\rho(r)/\rho(0)$. (D) Logarithm plot of normalized ED as a function of distance squares. (E) The extent of deviation from spherical symmetry. (F) Blurring ΔB-factor added to data for reducing the effective resolution.

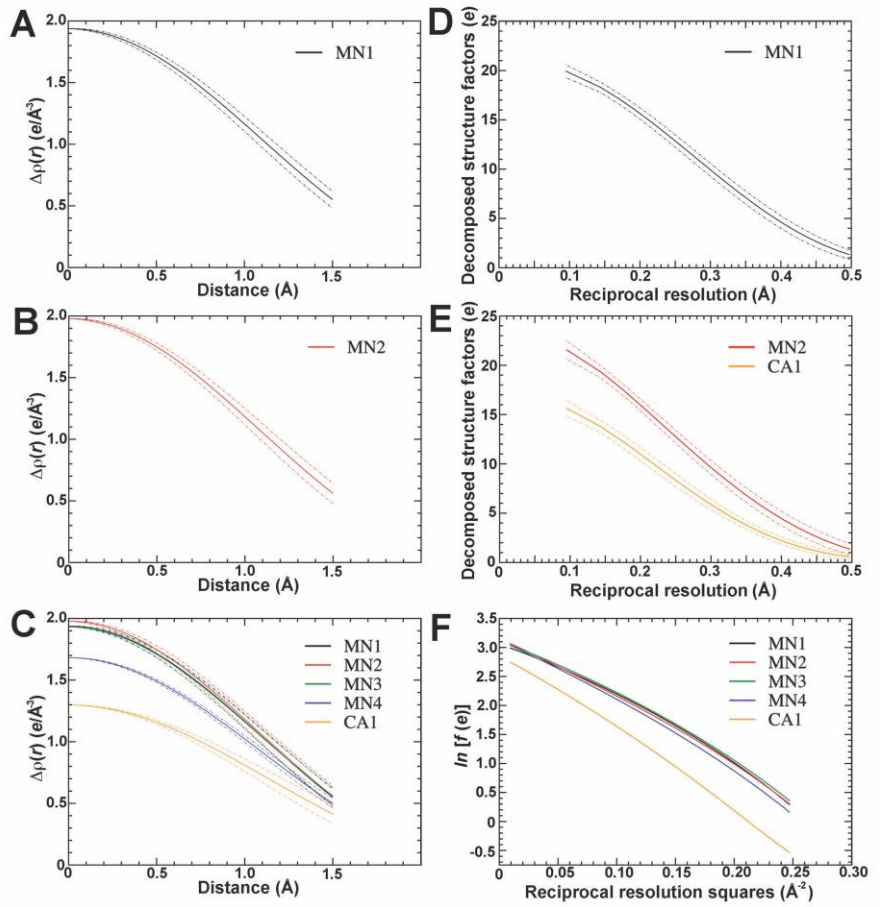

Figure S4. Deviation from spherical symmetry and decomposed structure factors for individual metal ions of the OEC in the 6jlj structure. (A-C) The mean values of spherical averaged ED peaks with ± one standard deviation of deviation from the spherical symmetry as a function of distance for Mn1 (A), Mn2 (B), and all metal ions (C). (D-F) Corresponding structure factors for individual metal ions of the OEC as a function of reciprocal resolution for Mn1 (D), Mn2 and Ca1 (E) or logarithm plot of spherically averaged structure factors as a function of reciprocal resolution squares for all metal ions of OEC (F).

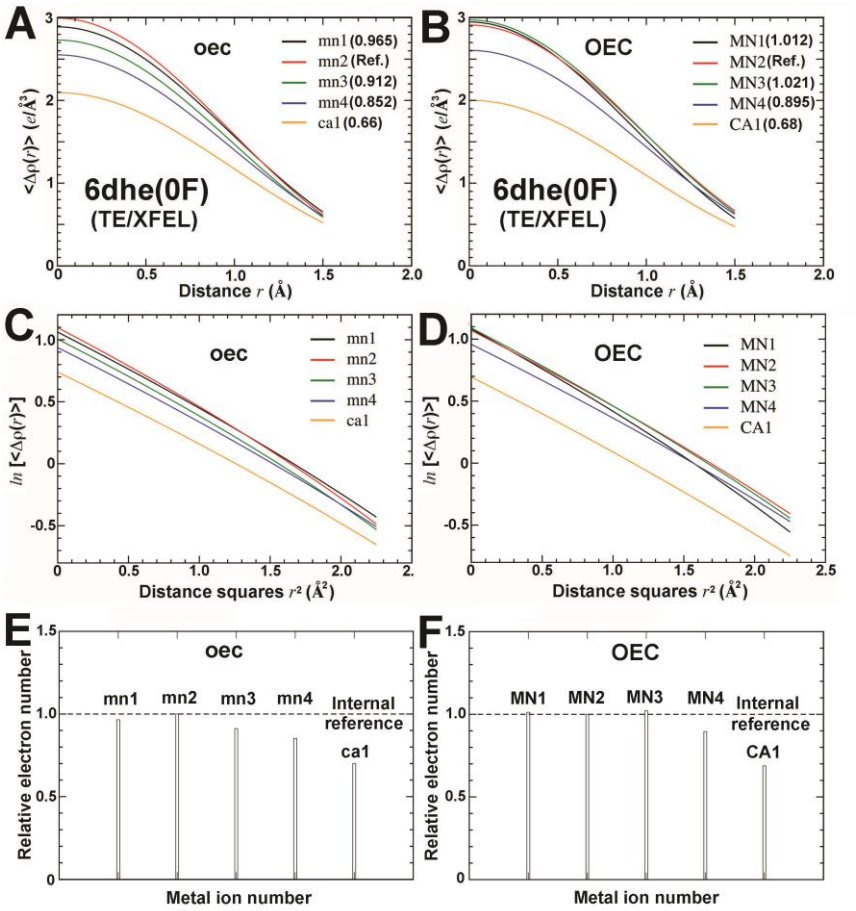

Figure S5. Density plots for the two OECs of the 6dhe structure. (A, B) Spherically averaged ED plot as a function of radial distance for the oec, and OEC. (C, D) Logarithm of ED as a function of radial distance squares for the oec and OEC. (E, F) Bar graph representation of relative ENs for Mn and Ca ions of the oec and OEC.

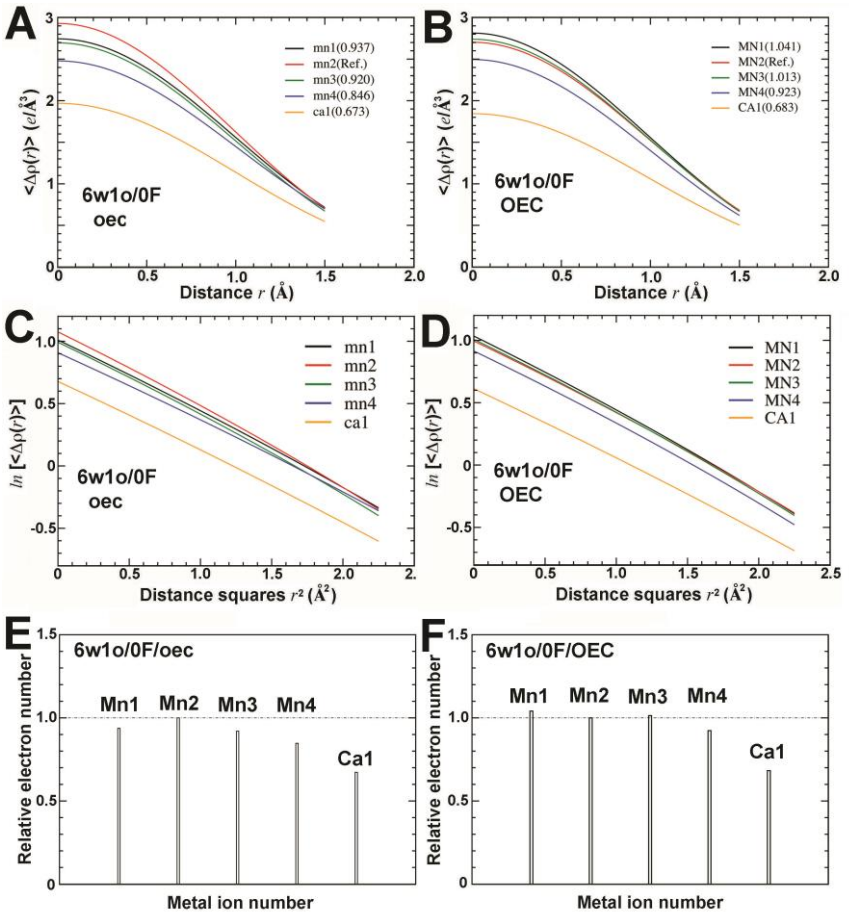

Figure S6. Density plots for the two OECs of the 6w1o structure. (A, B) Spherically averaged ED plot as a function of radial distance for the oec, and OEC. (C, D) Logarithm of ED as a function of radial distance squares for the oec and OEC. (E, F) Bar graph representation of relative ENs for Mn and Ca ions of the oec and OEC.

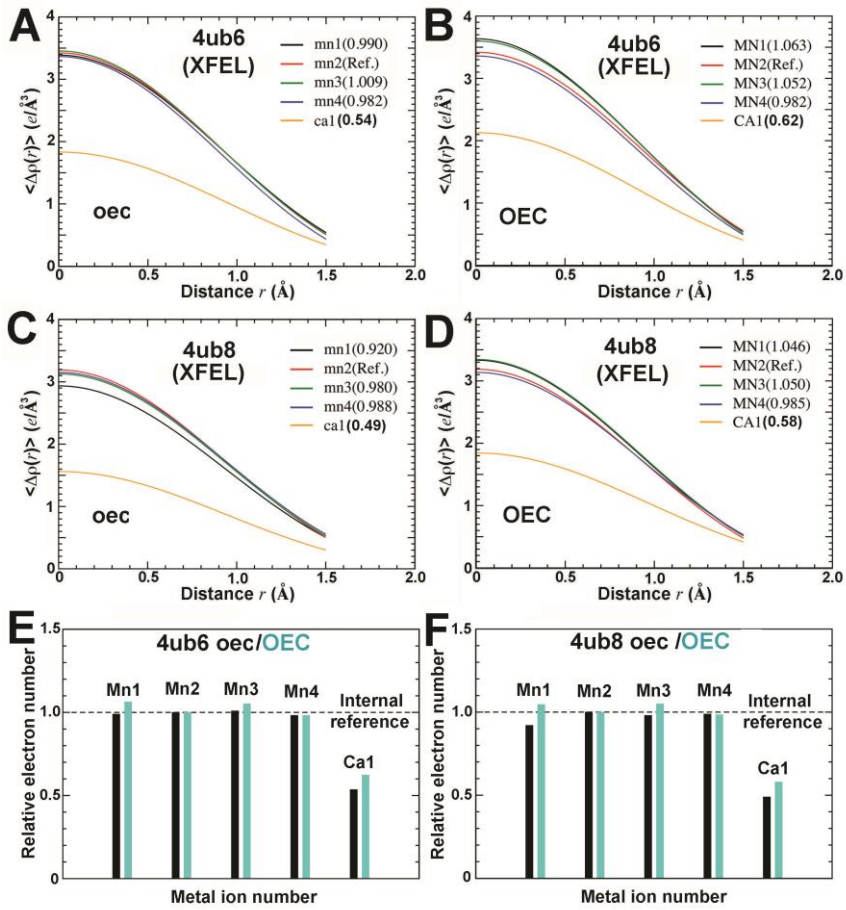

Figure S7. Density plots for the two OECs of the 4ub6 (A, B, E) and 4ub8 (C, D, F) structures. (A-D) Spherically averaged ED plot as a function of radial distance for the oec (A, C), and OEC (B, D). (E, F) Bar graph representation of relative ENs for Mn and Ca ions of the oec (black) and OEC (cyan).

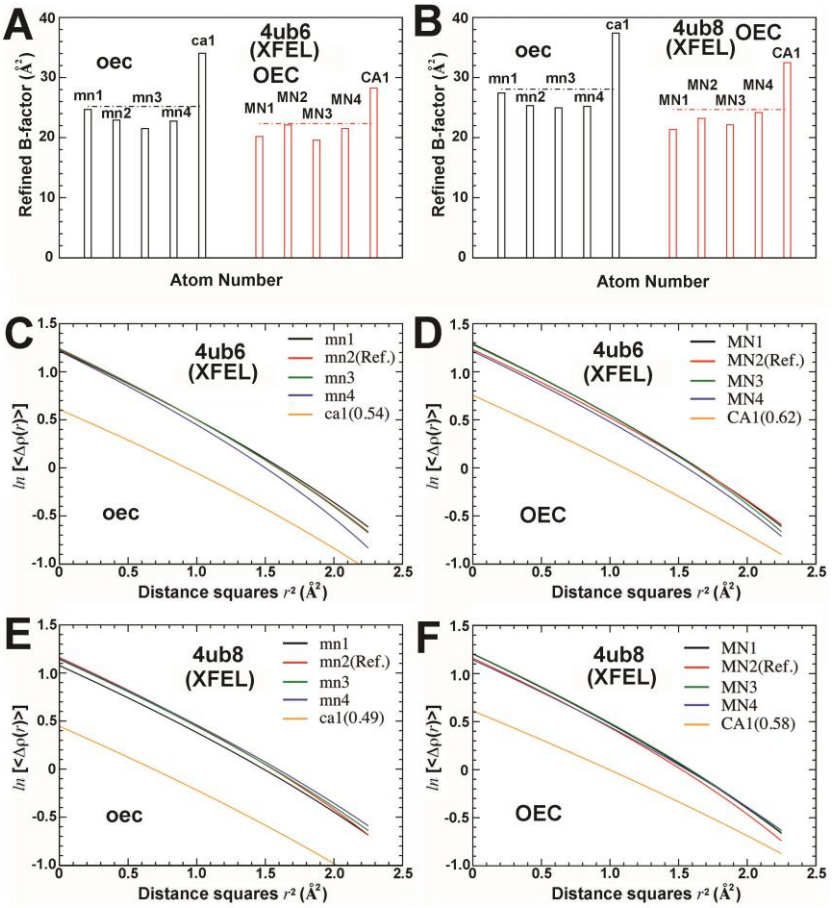

Figure S8. Refined B-factor and density plots for the two OECs of the 4ub6 (A, C, E) and 4ub8 (B, D, F) structures. (A, B) Bar graph representation of reported atomic B-factors for the oec (black) and OEC (red) in the 4ub6 (A), and 4ub8 (B) structures. (C-F) Logarithm of density as a function of radial distance squares for the oec (C, E) and OEC (D, F).