Predicting the oxidation states of Mn ions in the oxygen-evolving complex of photosystem II using supervised and unsupervised machine learning

Muhamed Amin

Received: 1 March 2022 / Accepted: 13 July 2022 / Published online: 27 July 2022
© The Author(s) 2022

Abstract
Serial Femtosecond Crystallography at the X-ray Free Electron Laser (XFEL) sources enabled the imaging of the catalytic intermediates of the oxygen evolution reaction of Photosystem II (PSII). However, due to the incoherent transition of the S-states, the resolved structures are a convolution from different catalytic states. Here, we train Decision Tree Classifier and K-means clustering models on Mn compounds obtained from the Cambridge Crystallographic Database to predict the S-state of the X-ray, XFEL, and CryoEM structures by predicting the Mn’s oxidation states in the oxygen-evolving complex. The model agrees mostly with the XFEL structures in the dark S1 state. However, significant discrepancies are observed for the excited XFEL states (S2, S3, and S0) and the dark states of the X-ray and CryoEM structures. Furthermore, there is a mismatch between the predicted S-states within the two monomers of the same dimer, mainly in the excited states. We validated our model against other metalloenzymes, the valence bond model and the Mn spin densities calculated using density functional theory for two of the mismatched predictions of PSII. The model suggests designing a more optimized sample delivery and illumination systems are crucial to precisely resolve the geometry of the advanced S-states to overcome the noncoherent S-state transition. In addition, significant radiation damage is observed in X-ray and CryoEM structures, particularly at the dangler Mn center (Mn4). Our model represents a valuable tool for investigating the electronic structure of the catalytic metal cluster of PSII to understand the water splitting mechanism.

Keywords Serial femtosecond crystallography · Oxygen-evolving complex · Machine learning

Introduction
Throughout history, Nature has inspired humans and driven many discoveries and inventions. For scientists, Nature is a vast school in which to observe, record, learn, and get inspired. Likewise, remarkable technologies and products have been inspired in one way or another by lessons learned from the environment, from sailing to flying to Velcro. Similarly, understanding the machinery of the biological nano-engines responsible for Photosynthesis would help understand how common metals such as manganese (Mn) act within Photosystem II (PSII) as the best catalyst for water oxidation (Fig. 1) (Brudvig 2008). A comprehensive understanding of the machinery can pave the way for developing similar artificial catalysts (Barber 2014; Nocera 2012, 2017; Orio and Pantazis 2021; Zhang and Reisner 2020). After recognizing the need to transition to renewable energy sources, Giacomo Ciamician proposed that photochemical devices can convert solar energy into fuel (Ciamician 1912). This idea was one of the early motivations to develop artificial Photosynthesis.

In natural Photosynthesis, PSII, a membrane protein complex in cyanobacteria, algae, and higher plants, harvests solar energy to drive water oxidization, converting the light energy into chemical energy and releasing di-molecular oxygen as a byproduct. In the ‘70 s of the twentieth century,
Bassel Kok described the biological water oxidation process in a five-step reaction (Kok et al. 1970). PSII carries out this reaction by coupling four-electron water oxidation at the oxygen-evolving complex (OEC), with the one-electron photochemistry occurring at the reaction center (Yano and Yachandra 2014; Vinyard and Brudvig 2017; Cox and Messinger 2013). The OEC consists of a heteronuclear Mn4O5Ca cluster, and it cycles through five intermediate S-states (S₀ to S₄) that corresponds to the abstraction of four successive electrons from the OEC (Kok et al. 1970). Several cofactors, i.e., chlorophyll, pheophytin, quinones, non-heme iron, and a redox-active tyrosine sidechain, are involved in the charge separation reaction during the water oxidation (Fig. 1).

A complete understanding of the catalytic activity of the PSII requires detailed information about the geometric and the electronic structure of the Mn₄CaO₅ cluster. The atomic geometric structure was first revealed at a resolution of 1.9 Å in 2011, using synchrotron X-ray crystallography (Umena et al. 2011a). However, synchrotron radiation induces sample damage, which result in systematic elongation in the metal–ligand bond distances (Garman 2010; Garman and Weik 2013; Grabolle et al. 2006). On the other hand, the recent advancement in the X-ray Free-electron Laser (XFEL) crystallography provided a radiation damage-free geometric structure of the Mn₄CaO₅ cluster (Suga et al. 2019, 2015, 2017a; Ibrahim et al. 2020; Kern et al. 2018a; Hussein et al. 2021). Moreover, XFEL studies provided the geometric structure for the S₁ state (dark-adapted) and other S-states at a resolution of ~ 2.0 Å, which shows a significantly reduced Mn-ligands bond lengths. However, despite the advancements in understanding the geometric structure of the Mn₄CaO₅ cluster, the electronic structure is still elusive.

Although the serial femtosecond crystallography at the XFEL sources provides a tool for imaging the catalytic intermediates of the Kok cycle, it is very challenging to state if the imaged structure is for the desired S-state. This is mainly because of the incoherent S-state transitions. In addition, a high resolution is required to accurately resolve the Mn-ligands positions because of the high density of Mn. Thus, we built machine learning models to predict the oxidation states of Mn in the OEC and hence the S-state of the

Fig. 1 The structure of the monomeric PSII with all the subunits is shown as a cartoon in gray. All the redox-active cofactors are involved in charge transfer are shown; the manganese ions of the OEC depicted in purple, the chlorophyll (Chl) in green, the pheophytin (Pheo) in yellow, the non-heme iron (Fe) in red, the quinones (Q) in light blue, and the tyrosine (Yₜ) in magenta. On the bottom right, Kok’s cycle of the oxygen evolution that takes place at the OEC of the PSII is shown. It shows the steps of the water oxidation reaction that is triggered by the absorption of photons shown as five oxidation states (S₀ → S₄)
X-ray, XFEL, and cryoEM structures. A similar method has been used previously to predict the metal oxidation states in metal–organic frameworks (Jablonska et al. 2021). The model is trained on Mn-containing small molecules obtained from the Cambridge Crystallographic Database (CCD), where the oxidation states are already known. The models, which showed very high accuracy scores (above 95%) on the training dataset, agreed mostly with the XFEL structures in the dark-adapted state \((S_1)\). However, significant discrepancies are observed for the X-ray and cryoEM \(S_1\) structures, as well as the illuminated XFEL structures \((S_2, S_3,\) and \(S_0)\). These disagreements might be the insufficient resolution or the incoherent S-state transitions of the OEC. The model is validated against another metalloenzymes and was tested against the calculated spin densities using density functional theory (DFT) of the Mn and the valence bond model in two PSII structures. Our model could be used to quickly evaluate structural models of the OEC. In addition, although other experimental techniques such as X-ray emission spectroscopy (XES) could be used to evaluate the total oxidation state of the cluster, they cannot assign oxidation state for each Mn center. Thus, the information provided by our model may be used to drive a possible mechanism for the catalytic reaction by monitoring the change in the oxidation state of each Mn center.

**Results and discussion**

To predict the oxidation states of the Mn in the oxygen evolution complex (OEC) of Photosystem II, we built a prediction model based on the data we collected for Mn compounds from the Cambridge Crystallography Database, where the oxidation states of the Mn are already known. Only small compounds with crystallographic data, R-factors \(\leq 0.075\), and are error-free (at the level of 0.05 Å) were included in the search (Bruno et al. 2002). Furthermore, only octahedral Mn compounds with oxygen (O) and nitrogen (N) ligands were selected because the Mn ions in the OEC are coordinated by O and N ligands. In total, the database that was built contains 1734, 835, and 107 structures corresponding to the oxidation states Mn(II), Mn(III), and Mn(IV), respectively.

The average bond lengths between the Mn and the ligands are significantly different for the different oxidation states; shorter for higher Mn oxidation states. Furthermore, in the case of Mn(III), the axial ligands have a significantly longer bond length due to the Jahn Teller effect. Therefore, the prediction model was designed to assess two features: (1) The average bond length between the Mn and the equatorial ligands (2) and the axial ligands. Figure 2a shows the average bond lengths of the equatorial ligands (X-axis) against the axial ligands (Y-axis). Although there are clear distinguished clusters for Mn(II) (Cyan), Mn(III) (Blue), and Mn(IV) (Dark Blue), there are some mislabeled elements within each cluster. There are several reasons for the mislabeled data, such as inter-ligand steric effects (Shields et al. 2000). The same conclusions were observed when calculating the oxidation states of metals using the valence bond model (Reeves et al. 2019). In addition, the asymmetry between the X and Y axes is observable in the Mn(III) cluster due to the Jahn Teller distortion.

The K-means clustering algorithm was used to label our data into three distinct clusters to correct the mislabeled
data. As shown in Fig. 2b, the algorithm converged, and the data were clustered into three clusters with the following centers: (2.18, 2.28) Å for Mn(II), (1.95, 2.26) Å for Mn(III), and (1.91, 2.05) Å for Mn(IV). The ratio Y/X for the three centers are (1.05, 1.16 1.08) Å, corresponding to Mn(II), Mn(III), and Mn(VI), respectively. The cluster that corresponds to Mn(III) clearly shows the effect of Jahn Teller distortion, where the axial ligands are significantly longer than the equatorial ones. Finally, we used the clustered data to build a prediction model based on two different classifiers: Gaussian Naïve Bayes and Decision Tree classifier.

**Gaussian Naïve Bayes classifier (GNB)**

The reason for choosing the GNB is that the data could be fitted to 2D Gaussians. Thus, we expected the model to perform well given the training dataset. The model is trained on 75% of the data, and the remaining 25% are used for testing. Before processing the data with K-means clustering, the accuracy score for the GNB prediction model is 94%, and the confusion matrix that shows the prediction against the true labels is shown in Fig. 3 (upper left). Furthermore, we performed a tenfold cross-validation to evaluate further the model, which resulted in a mean accuracy score of 96% and a sigma of 1%.

The accuracy score increased to 99% after using the processed data after the clustering, which is also reflected in the confusion matrix shown in Fig. 3 (upper right). Most of the wrong predictions are Mn(IV) data points, which were classified as Mn(III) (6 out of 44). Interestingly, the means of the Gaussians used to calculate the prior probabilities precisely match the center of the clusters obtained from the K-means algorithm.

**Decision tree classifier (DT)**

The Decision Tree (DT) classifier is based on a very different algorithm than the Naïve Bayes. Each node in the tree applies a test on a feature (here, the average bond lengths of the axial or the equatorial ligands); the branches descending from each node correspond to one of the possible values for that feature. The nodes are arranged in the tree so that the reduction in the information entropy is maximized.

The DT model shows a higher accuracy score than the GNB before and after the clustering (Fig. 3 lower left, lower right, respectively). The cross-validation calculations show an accuracy score of 95% before the clustering and 100% after the clustering. The set of rules used to classify the Mn oxidation states are shown in Fig. 4.

**Prediction of the catalytic states of the Kok cycle**

The OEC contains four Mn ions and one Ca ion (Fig. 1). The Mn ions are ligated mainly by (O) and one (N). During the catalytic cycle of the PSII enzyme, the Mn ions are oxidized in the transition between the S-states. According to different spectroscopic studies, using different techniques, the oxidation state of the Mn ions in the dark-adapted state of the OEC (S1-state) is Mn(III, IV, IV, III) (Visser et al. 2002; Riggs et al. 1992; Bergmann et al. 1998; Cox et al. 2014). The Mn(III) is then oxidized in the transition to higher S-states till all Mn are Mn(IV) in the S3-state (Fig. 1). Our prediction model with the highest accuracy score based on Decision Tree Classifier was used to predict the oxidation states of the Mn in Photosystem II in 38 structures (27 XFEL, 6 X-ray, and 5 cryoEM structures) for each monomer independently. The Decision Tree Classifier trained on data prior clustering is also tested, which predicted a slightly more reduced structure (see supporting information). Only the structures of the

Fig. 3 The confusion matrix of GNB (top) and DT (bottom) prediction models before (left) and after (right) k-means clustering

Fig. 4 The fitted Decision Tree. E and A are the average bond lengths of the equatorial and axial ligands, respectively. The left branches are followed if the condition is satisfied
meta-stables S-states (S₁, S₂, S₃, and S₀), which have a resolution of 2.5 Å or better, were included in this study (Table 1). After the predictions of the oxidation states, the S-states were assigned according to the total charges of the 4 Mn, i.e., the total charges for S₀, S₁, S₂, and S₃ are 13, 14, 15, and 16, respectively. In addition, we have included the S−₁, S−₂, ..., S−₅ to account for more reduced structures.

The prediction accuracy of our model is 96% before clustering and nearly 100% after clustering when applied to the small molecules. To further assess our model, we predicted the oxidation states of the Mn in DFT optimized structures of the OEC, where the oxidation states are already known from the spin densities (Amin et al. 2019). The model successfully predicted the oxidation state that matches the Mn’s

| PDBID          | S-state Monomer 1 | S-state Monomer 2 | Method     |
|----------------|-------------------|-------------------|------------|
| 3wu2 (Umena et al. 2011b) | S₁, S₂ | Mn(III, III, III) | X-ray      |
| 4if6 (Koua et al. 2013)   | S₁, S₀ | Mn(III, IV, III) | X-ray      |
| 4pj6 (Hellmich et al. 2014) | S₁, S⁻₁ | Mn(III, IV, III) | X-ray      |
| 4ub6 (Suga et al. 2015)   | S₁, S₀ | Mn(III, IV, III) | XFEL       |
| 4ub8 (Suga et al. 2015)   | S₁, S₀ | Mn(III, IV, III) | XFEL       |
| 5bSe (Tanaka et al. 2017) | S₁, S₀ | Mn(III, IV, III) | X-ray      |
| 5b66 (Tanaka et al. 2017) | S₁, S₀ | Mn(III, IV, III) | X-ray      |
| 5gh (Suga et al. 2017)    | S₁, S₀ | Mn(III, IV, III) | XFEL       |
| 5gi (Suga et al. 2017)    | S₁, S₀ | Mn(III, IV, III) | XFEL       |
| 5h2 (Uto et al. 2017)     | S₁, S⁻₁ | Mn(III, III, III) | X-ray      |
| 5is (Young et al. 2016)   | S₁, S₀ | Mn(III, IV, III) | XFEL       |
| 5ws5 (Suga et al. 2017)   | S₁, S₀ | Mn(III, IV, III) | XFEL       |
| 5ws6 (Suga et al. 2017)   | S₁, S₀ | Mn(III, IV, III) | XFEL       |
| 5zzn (Nakajima et al. 2018) | S₁, S⁻₁ | Mn(III, IV, III) | XFEL       |
| 6dhe (Kern et al. 2018b)  | S₁, S₀ | Mn(III, IV, III) | XFEL       |
| 6dhf (Kern et al. 2018b)  | S₁, S₀ | Mn(III, IV, III) | XFEL       |
| 6dho (Kern et al. 2018b)  | S₁, S₀ | Mn(III, IV, III) | XFEL       |
| 6jij (Suga et al. 2019)   | S₁, S₀ | Mn(III, IV, III) | XFEL       |
| 6jik (Suga et al. 2019)   | S₁, S₀ | Mn(III, IV, III) | XFEL       |
| 6jil (Suga et al. 2019)   | S₁, S₀ | Mn(III, IV, III) | XFEL       |
| 6jin (Suga et al. 2019)   | S₁, S₀ | Mn(III, IV, III) | XFEL       |
| 6jio (Suga et al. 2019)   | S₁, S₀ | Mn(III, IV, III) | XFEL       |
| 6jip (Suga et al. 2019)   | S₁, S₀ | Mn(III, IV, III) | XFEL       |
| 6wj1 (Ibrahim et al. 2020) | S₁, S₀ | Mn(III, IV, III) | XFEL       |
| 6wj2 (Ibrahim et al. 2020) | S₁, S₀ | Mn(III, IV, III) | XFEL       |
| 7cji (Li et al. 2021)     | S₁, S₀ | Mn(III, IV, III) | XFEL       |
| 7cjj (Li et al. 2021)     | S₁, S₀ | Mn(III, IV, III) | XFEL       |
| 7cou (Li et al. 2021)     | S₁, S₀ | Mn(III, IV, III) | XFEL       |
| 7di1 (Kato et al. 2021)   | S₁, S⁻₄ | Mn(III, III, II) | XFEL       |
| 7dt (Kato et al. 2021)    | S₁, S⁻₂ | Mn(III, III, II) | XFEL       |
| 7f80 (Gisriel et al. 2022) | S₁, S⁻₅ | Mn(III, III, II) | XFEL       |
| 7rcv (Gisriel et al. 2022) | S₁, S⁻₅ | Mn(III, III, II) | XFEL       |
| 7rf2 (Hussein et al. 2021) | S₁, S₀ | Mn(III, IV, IV) | XFEL       |
| 7rf3 (Hussein et al. 2021) | S₁, S₀ | Mn(III, IV, IV) | XFEL       |
| 7r88 (Hussein et al. 2021) | S₁, S₀ | Mn(III, IV, IV) | XFEL       |

The PDB IDs are sorted alphabetically. The bolded entries are for the structures that the predicted S-state matched the reported
spin densities. On the other hand, the first impression of the prediction model’s accuracy to predict the oxidation states of the OEC is low. As shown in Fig. 5, the structures are mostly in the S1 state for both monomers. However, the experimentally assigned S-states, which are based on the number of flashes used to pump the sample, are significantly different from the predicted by our model. For monomer 1, out of the 38 structures, the prediction matched the experimentally assigned S-state for ten structures; all of them are assigned as S1 except for the 6dhf and 6w1v structures, which are S2 and S3, respectively (Table 1). For monomer 2, 10 structures matched the experimentally assigned S-state, all of them for S1 except for the 7rf3 structure, which is S2. (PDB ID: 6dhp). The mismatch between the two monomers may be attributed to the different turnover rates in the crystal due to the different physical/chemical conditions in the crystals (Wang et al. 2021).

A closer and more detailed look at the predicted oxidation states shows a totally different story. Among the investigated PSII PDB files, 22 structures correspond to the S1-state (PDB; 3wu2, 4il6, 4pj0, 4ub6, 4ub8, 5b66, 5b5e, 5gth, 5h2f, 5sw5, 5zzn, 6dhe, 6lj1, 6ljm, 6w1o, 7cji, 7cou, 7rf3, 7d1t, 7d1u, 7n8o, 7rcv); eleven out of the 22 have been solved using XFEL data. In seven XFEL-S1 structures, the predicted S-states, using our models, agreed with the experimental one in both monomers. Furthermore, in another three XFEL-S1 structures, the predicted S-state agreed with the experimental one in at least one of the two monomers. Although, the predicted S-state did not match the reported one in only one XFEL-S1 structure. The agreement between the XFEL structures and the prediction in the dark state (S1) supports that our models accurately predict the damage-free XFEL-S1 structures.

The other 11 structures were solved using data that have been collected using conventional synchrotron X-ray radiation (PDB; 5zzn, 7n8o, 7rcv, 7d1t, 7d1u). While the predicted and reports states are mostly in agreement, for the XFEL S1-structure, the reported S-states for these 11 structures mainly disagreed with the predicted one. Only one monomer of the 11 structures was predicted to be in the S1-state as reported, while the rest are predicted to be in more reduced states S0, S−1, …, or S−5 (Table 1). Some of these S1-models were investigated theoretically, and it was predicted to suffer from severe radiation damage (Luber et al. 2011; Kato et al. 2021). It is worth noticing that the single agreement (in the case of the X-ray S1 structures) is coming from a collected dataset using a low dose of synchrotron radiation (Suga et al. 2015). These observations emphasize the influence of the radiation damage on the OEC electronic structure and hence the geometrical structure, in agreement with several studies that generally assess the radiation damage in protein crystallography (Garman 2010; Garman and Weik 2013, 2011; Grabolle et al. 2006; Garman and McSweeney 2007; Garman and Nave 2009; Hendrickson 1991). Moreover, several studies have discussed the radiation damage in the case of PSII,

![Fig. 5](image-url) A histogram of the predicted for monomer1 and monomer and the labeled S-states of the 34 structures of the OEC. The S-states are assigned according to the total charges of the 4 Mn in the OEC
particularly during X-ray (Askerka et al. 2015; Yano et al. 2005), and the cryoEM data collection (Kato et al. 2021).

On the other hand, the agreement of the prediction was relatively low for the XFEL structures (PDB: 5gti, 5tis, 5ws6, 6dhf, 6dho, 6dhp, 6jlk, 6jll, 6jlo, 6jlp, 6w1p, 6w1v, 7cjij, 7rfl3, 7rfl8), which are for different excited S-states, S2, S3, and S0. Although Mn emission spectroscopy shows a shift in the spectrum due to the Mn oxidation in the illuminated structures compared to the dark state, the shift may result from a low population of the excited S-states. Furthermore, even if a coherent transition for all nanocrystals takes place, the resolution may not be enough to accurately resolve the Mn-ligand positions, which may explain the discrepancy between the assigned and predicted S-states. In addition, we cannot eliminate the possibility that the presence of the Ca ion in the OEC affects the cluster’s geometry during the light activation. The Ca ion in the OEC has been intensively investigated; it plays a critical role in the substrate insertion during the light activation (Boussac et al. 2004; Cox et al. 2011; Koua et al. 2013).

Overall, the predicted oxidation state of the Mn ions for the XFEL structures did not mostly show Mn(II) content, except for PDB 7cou, 7cji, 6jlp, where one of the two monomers contained a Mn ion that was predicted to be in Mn(II) oxidation state. Unlike the XFEL structures, most of the synchrotron or cryoEM structures contain Mn(II), and all Mn in the 7n8o, 7rcv structures are reduced to Mn(II), likely because of the high radiation dose (Gisriel et al. 2022). In addition, these structures show higher reductions states, i.e., S−1, ..., or S−5, that are physiologically do not exist in an active PSII, indicating that the suffering of radiation damage significantly influences the electronic and geometrical structure of the OEC. The radiation damage influences the OEC geometry, and it is manifested in the prediction results of the S1-structures. These results emphasize the importance of radiation-free data collection to study a functional PSII.

Interestingly, for all the structures that showed a presence of Mn(II) in the OEC, the Mn(II) oxidation state was permanently assigned to the Mn4 (the dangling Mn), indicating the high vulnerability of this Mn ion. It is recently suggested that Mn4 is a Mn(II) high-affinity site, where the Mn ions are oxidized in preparation for the OEC formation during the OEC assembly (Mino and Asada 2021).

According to our model, Mn2, Mn3 are mostly in the Mn(IV) oxidation state, while Mn1 and Mn4 is mostly in the Mn(III) states (Fig. 6), which agrees with several theoretical studies proposed that Mn2 and Mn3 are oxidized in the S1 state. The transition to S2 takes place by likely oxidizing either Mn4 (as in 6dhf) or Mn1 (as in 6dho). According to theoretical studies supported by EPR measurements, the oxidation of Mn4 in the S2 is responsible for the g = 2 EPR signal, while the multi-lines g = 4.1 signal may be attributed to the oxidation of Mn1 or the conversion of O4 from a µ-oxo to hydroxo bridge (Corry and O’Malley 2019). Furthermore, several studies suggested that the transition between the S1 and the S3 will involve the oxidation of Mn4, which is then reduced and Mn1 is oxidized before both Mn are oxidized in the S3 state. It is interesting to notice that this sequence of the reaction is observed in the structures resolved by Kern et. al. (6dhe, 6dhf, 6dho, 6dhp) (2018b) which are predicted to be in the S1, S2,Mn4(IV), S2,Mn1(IV) and S3 states (Table 1) (Amin et al. 2019; Marius Retegan et al. 2016; Kaur et al.)

![Fig 6](image-url) The distribution of the oxidation states of the 4 Mn ions in the OEC (Mn(II): cyan, Mn(III): blue, Mn(IV): dark blue) for monomers 1 and 2.
Although, the 6dhp structure is assigned for S₃, the existence of the additional water ligand near Mn1 support the prediction by our model.

**Model validation**

The OEC is a unique catalytic cluster, which has not been synthesized in artificial system. Thus, the transferreblity of the machine learning model, which is trained on a small compounds is questionable. However, since the model is predicting the oxidation state of each Mn center invidually based on its coordination chemistry and does not predict the chemical properties of the cluster, the predictions should be chemically reasonable. This assumption is valid because our data sets contains more than 1200 structures with single Mn, which are placed randomly in the training and test sets in tenfold cross validation steps and produced very high accuracy scores. In other words, the structures with mono-Mn center produces high accuracy when predicting the oxidation states of Mn in structures with multi-Mn centers.

To further validate our model we used three independent methods: (1) We predicted the Mn oxidation states in other proteins structures with significantly higher resolution. (2) The OEC is a unique cluster. Thus, we calculated the spin densities of the Mn in two of the mismatched structures of different S-states resolved by different groups. (3) We validated our predictions against the valence bond model.

Prediction of Mn oxidation states in high resolution structures of Mn-superoxide dismutase and oxidase. The structure of Manganese Superoxide Dismutase from Sphingobacterium is solved at 1.35 Å resolution (PDB code: 5A9G). The Mn ions in the two dimers are in the Mn(II) oxidation state. The averages of the axial and equatorial ligand bond lengths are 2.2 Å and 2.1 Å, respectively, in both monomers (unlike the PSII structures, where the two monomer are significantly different). It is clear that the Mn ions belong to the Mn(II) cluster (Fig. 2b) and the model predicts the correct oxidation state for Mn in the two monomers. In addition, we predicted the oxidation state of Mn in the crystal structure of R2-like ligand-binding oxidase from Saccharopolyspora Erythraea solved at 1.38 Å resolution. The Mn is assigned Mn(III) oxidation state, which is predicted correctly by our model. The averages of the axial and equatorial ligand bond lengths are 2.3 and 2.0 Å, respectively. The Jahn Teller distortion is clearly observed and the Mn belongs to the cluster of Mn(III) in Fig. 2b.

Comparing the predicted Mn oxidation states in the OEC against the calculated Mn spin densities from DFT. Because the OEC is a unique catalytic center, we validated our predictions against the calculated Mn spin densities from DFT. We calculated the spin densities for the Mn centers in two mismatching structures obtained from different groups for advanced S-states (since a good agreement is obtained for structures in the dark state) using Gaussian09 with B3LYP/6-31G(d) level of theory. The first structure is the 6dho, which is assigned a S₃ state, while it was predicted to be in the S₂ by the model. The second is 6jlk, which is assigned the S₂ state and predicted to be in the S₁ state.

The calculated Mn spin densities in the 6dho show that Mn4 is the most reduced center (Table 2), which is predicted to be in the Mn(III) by our model. Furthermore, after the optimizations all Mn are predicted to be in the Mn(IV) state, which agrees with the DFT spin densities. Similarly, the spin densities of the Mn in the 6jlk structure suggest that the appropriate state for this structure is S₁, which agrees with our model (Table 2). However, after the optimization, the

| Table 2 | The Mn spin densities in the XFEL and DFT optimized structures of the OEC compared to the machine learning and the valence bond models |
|---------|---------------------------------------------------------------------------------------------------------------|
|         | 6dho                                                                                                         | 6jlk                                                                                           |
|         | The structure is assigned S₃ and predicted to be in the S₃ state                                             | The structure is assigned S₂ and predicted to be in the S₂ state                               |
| Spin densities |                                                                                                              |                                                                                           |
| Mn1    | 3.12                                                             | 3.76                                                             |
| Mn2    | 2.94                                                             | 2.99                                                             |
| Mn3    | 3.24                                                             | 3.26                                                             |
| Mn4    | 3.49                                                             | 4.14                                                             |
| Prediction |                                                                                                              |                                                                                           |
| Mn1    | IV                                                               | III                                                            |
| Mn2    | IV                                                               | IV                                                             |
| Mn3    | IV                                                               | III                                                            |
| Mn4    | III                                                              | III                                                            |
| Valence bond model |                                                                                                              |                                                                                           |
| Mn1    | IV                                                               | III                                                            |
| Mn2    | IV                                                               | III                                                            |
| Mn3    | IV                                                               | III                                                            |
| Mn4    | III                                                              | III                                                            |
| Valence bond parameterized based on DFT/EXAFs model. (Laber et al. 2011) |                                                                                           |
| Optimized 6dho |                                                                                                              | Optimized 6jlk                                                                                 |
| Spin densities |                                                                                                              |                                                                                           |
| Mn1    | 2.98                                                             | 3.87                                                             |
| Mn2    | 2.98                                                             | 3.02                                                             |
| Mn3    | 2.94                                                             | 2.95                                                             |
| Mn4    | 2.89                                                             | 2.92                                                             |
| Prediction |                                                                                                              |                                                                                           |
| Mn1    | IV                                                               | III                                                            |
| Mn2    | IV                                                               | IV                                                             |
| Mn3    | IV                                                               | IV                                                             |
| Mn4    | IV                                                               | IV                                                             |
| Valence bond model |                                                                                                              |                                                                                           |
| Mn1    | IV                                                               | III                                                            |
| Mn2    | IV                                                               | III                                                            |
| Mn3    | IV                                                               | III                                                            |
| Mn4    | IV                                                               | IV                                                             |
| Valence bond parameterized based on DFT/EXAFs model. (Laber et al. 2011) |                                                                                           |

The prediction matches well with the calculated spin densities before and after the DFT geometry optimizations.

Springer
DFT spin densities and our model confirm that the structure is at the S_2 state.

Valence Bond Model (VBM). According to the valence bond model the oxidation state of the metal center could be calculated as the sum of valence from each bond:

\[ v = \sum_i c_i \left( \frac{R_i - R_0}{B} \right), \]  

(1)

where \( R_0 \) and \( B \) are empirical parameters obtained from the IUCR dataset by David Brown and \( R_i \) is the bond length of the individual ligands (Reeves et al. 2019; Brown 2009, 2017). Using this model (Eq. 1) we calculated the oxidation states of the Mn’s in the 6dho and 6jlk structures (Table 2).

In general, the calculated oxidation states using VBM are more reduced than the predicted with the ML model. However, the calculations matches the ML model and the Mn spin densities for the DFT optimized structures.

Furthermore, we used the DFT structure of the S_1 state that showed a good match with the EXAFS (Luber et al. 2011) to estimate the empirical parameters \( R_0 \) in Eq. 1 and recalculated the oxidation states of Mn in the later structures. The calculated oxidation states based on the updated parameters agrees with the ML model for all Mn’s of all structures (prior and after the DFT optimizations) Table 2.

In addition to the previous validation steps, the model agrees mostly with the XFEL structures in the dark-S1 state, which provide another supporting evidence for the validity of the model.

**Conclusion**

In conclusion, we built two models based on the Decision Tree Classifier (DT) and the Gaussian Naïve Bayes Classifier (GNB) to predict the oxidation state of the Mn ions in the OEC using the available small molecules from the Cambridge Structure Data. The DT model showed better results than the GNB model; it has an accuracy of nearly 100% in the prediction of small molecules and ~75% in the case of XFEL-S_1 structures. Furthermore, the prediction of the synchrotron and cryoEM S_1 structures predicted reduced structures (S_{\text{red}}, S_{\text{S1}}, \ldots, S_{\text{S5}}), indicating the presence of severe radiation damage, in agreement with several studies that suggested the presence of radiation damage during data collection. The cryoEM structures, in particular, showed significantly high reduced states, up to S_{\text{S5}}. The observation of radiation damage signs emphasizes the importance of radiation-free data collection to investigate the functional OEC of PSII. In addition, the model predicted that Mn1 and Mn4 are more likely to be oxidized during the transitions S_1 \rightarrow S_2 and S_2 \rightarrow S_3 states. Moreover, the prediction model shows that Mn4 is the most susceptible Mn ion among the four ions to radiation damage.

Although the experimental methods such as XES or XANES are used to determine the oxidation states of the Mn_5O_2Ca^{2+} cluster, it cannot assign the oxidation states of each Mn separately, which is important to understand the mechanism of the water splitting reaction. Our model provides a tool for quickly evaluating the structure and to provide the oxidation states of each Mn center for the different structures of the S-states. Furthermore, the model could be used to evaluate the radiation damage in the X-ray and cryoEM structures.

**Methods**

**Data collection**

We used ConQuest software to search the Cambridge Structural Database for small molecules that contain Mn ions. The first quest resulted in more than 15,000 small molecules; however, several filters were used to end up with a reliable data set that resembles some features of the OEC. We started by eliminating the noncrystallographic structures, also any structure with R-factors ≥ 0.075. Another filter was added to improve the preciseness of the bond length in the structures by including only the error-free structures (at the level of 0.05 Å) (Bruno et al. 2002). Furthermore, only Mn compounds with oxygen (O) and nitrogen (N) ligands were included. Finally, overall, we built a database of thousands of octahedral coordination compounds containing 1734, 835, and 107 structures corresponding to the oxidation states Mn(II), Mn(III), and Mn(IV), respectively. The data includes 2795 structures that includes µ-oxo-bridges.

**Machine learning models**

Initially we used sklearn (Pedregosa et al. 2011) to build prediction models using Gaussian Naïve Bayes and Decision Tree classifiers based on several features: (1) each Mn-Ligand bond is a feature (i.e., six features) (2) The type of each atom ligating the Mn (i.e., six features). However, because both Mn oxidation state and the type of the ligand affect the Mn-Ligand distances, the same accuracy score is achieved with only two features: (1) the average distances from Mn to the equatorial and (2) axial ligands.

In-house python scripts are written to read the PDB files, extract the octahedral Mn ions, and calculate the distances between the Mn and the ligands using the BioPython (Cock et al. 2009) package. Then, a CSV file that contains the extracted data from all PDB files is created. Pandas library is used to parse the input data to the machine learning models. The K-means clustering algorithm in sklearn is used
to cluster the Mn ions based on the average bond length of the equatorial and axial ligands into three clusters. Each cluster represents a different oxidation state of the Mn; then, the supervised learning process based on Gaussian Naive Bayes and Decision Tree classifiers is repeated based on the clustered data. The split into training and test datasets is done randomly 10 folds by skewen with 70% of the data for training and 30% for testing using cross-validation function.

**Supplementary Information** The online version contains supplementary material available at [https://doi.org/10.1007/s11120-022-00941-8](https://doi.org/10.1007/s11120-022-00941-8).

**Acknowledgements** We thank Dr. Mohamed Ibrahim, Prof. Dr. Christian Limberg and Dr. Beatrice Cula (Institute of Chemistry, Humboldt Universität zu Berlin) for supporting the CSD data collection. We acknowledge the support from the DOE Grants DESC0001423 (M.R.G. and V.S.B.). We thank Prof. Victor Batista, Prof. Gary Brudvig, Prof. Marilyn Gunner and Dr. Jimin Wang for the useful discussion.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit [http://creativecommons.org/licenses/by/4.0/](http://creativecommons.org/licenses/by/4.0/).

**References**

Amin M, Kaur D, Yang KR, Wang J, Mohamed Z, Brudvig GW, Gunner MR, Batista V (2019) Thermodynamics of the S2-to-S3 state transition of the oxygen-evolving complex of photosystem II. Phys Chem Chem Phys 21(37):20840–20848

Askeresa M, Vinyard DJ, Wang J, Brudvig GW, Batista VS (2015) Analysis of the radiation-damage-free X-ray structure of Photosystem II in light of EXAFS and QM/MM data. Biochemistry 54(9):1713–1716

Barber J (2014) Photosystem II: its function, structure, and implications for artificial photosynthesis. Biochem Soc Trans 42(2):8350–8352

Boussac A, Rappaport F, Carrier P, Verbavatz J-M, Gobin R, Kirilovsky D, Rutherford AW, Sugiura M (2004) Biosynthetic Ca2+/Sr2+ substitution on the electronic structure of the oxygen-evolving complex of photosystem II from the mesophilic cyanobacterium, *Synechocystis sp.* PCC 6803. Proc Natl Acad Sci USA 119(1):e2116765118

Grabolle M, Haumann M, Müller C, Liebisch P, Dau H (2006) Rapid loss of structural motifs in the manganese complex of oxygenic photosynthesis by X-ray irradiation at 10–300 K. J Biol Chem 281(8):4580–4588

Hellmich J, Bonner M, Burkhardt A, Ibrahim M, Kern J, Meents A, Muh F, Dobbeck H, Zouni A (2014) Native-like photosystem II superstructure at 2.44 Å resolution through detergent extraction from the protein crystal. Structure 22(11):1607–1615

Hendrickson WA (1991) Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. Science 254:5028

Hussein R, Ibrahim M, Bhowmick A, Simon PS, Chatterjee R, Lassalle L, Doyle M, Bogacz I, Kim I-S, Cheah MH, Gul S, de Lichtenberg C, Chernev P, Pham CC, Young ID, Carbajo S, Fuller FD, Alonso-Mori R, Batyuk A, Sutherland KD, Brewster AS, Bolotovskiy R, Mendez D, Holton JM, Moriarty NW, Adams PD, Bergmann U, Sauter NK, Dobbeck H, Messinger J, Zouni A, Kern J, Yachandra VK, Yano J (2021) Structural dynamics in the water and proton channels of photosystem II during the S2 to S3 transition. Nat Commun 12(1):6531

Bruno JJ, Cole JC, Edgington PR, Kessler M, Macrae CF, McCabe P, Pearson J, Taylor R (2002) New software for searching the Cambridge Structural Database and visualizing crystal structures. Acta Crystallogr Sect B Struct Sci 58(3):389–397

Ciamician G (1912) The photochemistry of the future. Science 36(926):385–394

Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJ (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25(11):1422–1423

Corry TA, O’Malley PJ (2019) Proton isoformers rationalize the high- and low-spin forms of the S2 state intermediate in the water-oxidizing reaction of photosystem II. J Phys Chem Lett 10(7):5226–5230

Cox N, Messinger J (2013) Reflections on substrate water and dioxygen formation. Biochem Biophys Acta 1827(8–9):1020–1030

Cox N, Rapatskiy L, Su J-H, Pantazis DA, Sugiura M, Kulik L, Dorlet P, Rutherford AW, Neese F, Boussac A (2011) Effect of Ca2+/Sr2+ substitution on the electronic structure of the oxygen-evolving complex of photosystem II: a combined multiplicity EPR, 55Mn-ENDOR, and DFT study of the S2 state. J Am Chem Soc 133(10):3635–3648

Cox N, Retegan M, Neese F, Pantazis DA, Boussac A, Lubitz W (2014) Electronic structure of the oxygen-evolving complex in photosystem II prior to O2 bond formation. Science 345(6198):804–808

Garman EF (2010) Radiation damage in macromolecular crystallography: what is it and why should we care? Acta Crystallogr D Biol Crystallogr 66(4):339–351

Garman EF, McSweeney SM (2007) Progress in research into radiation damage in cryo-cooled macromolecular crystals. J Synchrotron Radiat 14(1):1–3

Garman EF, Nave C (2009) Radiation damage in protein crystals examined under various conditions by different methods. J Synchrotron Radiat 16(2):129–132

Garman EF, Weik M (2011) Macromolecular crystallography radiation damage research: what’s new? J Synchrotron Radiat 18(3):313–317

Garman EF, Weik M (2013) Radiation damage to biological macromolecules: some answers and more questions. J Synchrotron Radiat 20(1):1–6

Gijsfel C, Wang J, Liu J, Flesher DA, Reiss KM, Huang HL, Yang KR, Armstrong WH, Gunner MR, Batista VS, Debus RJ, Brudvig GW (2022) High-resolution cryo-electron microscopy structure of photosystem II from the mesophilic cyanobacterium, *Synechocystis* sp. PCC 6803. Proc Natl Acad Sci USA 119(1):e2116765118

Garbolke M, Haumann M, Müller C, Liebisch P, Dau H (2006) Rapid loss of structural motifs in the manganese complex of oxygenic photosynthesis by X-ray irradiation at 10–300 K. J Biol Chem 281(8):4580–4588

Hellmich J, Bonmer M, Burkhardt A, Ibrahim M, Kern J, Meents A, Muh F, Dobbeck H, Zouni A (2014) Native-like photosystem II superstructure at 2.44 Å resolution through detergent extraction from the protein crystal. Structure 22(11):1607–1615

Hendrickson WA (1991) Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. Science 254:5028

Hussein R, Ibrahim M, Bhowmick A, Simon PS, Chatterjee R, Lassalle L, Doyle M, Bogacz I, Kim I-S, Cheah MH, Gul S, de Lichtenberg C, Chernev P, Pham CC, Young ID, Carbajo S, Fuller FD, Alonso-Mori R, Batyuk A, Sutherland KD, Brewster AS, Bolotovskiy R, Mendez D, Holton JM, Moriarty NW, Adams PD, Bergmann U, Sauter NK, Dobbeck H, Messinger J, Zouni A, Kern J, Yachandra VK, Yano J (2021) Structural dynamics in the water and proton channels of photosystem II during the S2 to S3 transition. Nat Commun 12(1):6531
Ibrahim M, Fransson T, Chatterjee R, Cheah MH, Hussein R, Lassalle L, Sutherland KD, Young ID, Fuller FD, Gul S, Kim JS, Simon PS, de Lichtenberg C, Chernev P, Bogacz I, Pham CC, Orville AM, Saichek N, Northen T, Batyuk A, Carbajo S, Alonso-Mori R, Tono K, Owada S, Bhowmick A, Botolovsky R, Mendez D, Moriarty NW, Holton JM, Dobbek H, Brewster AS, Adams PD, Sauter NK, Bergmann U, Zouni A, Messinger J, Kern J, Yachandra VK, Yano J (2020) Untangling the sequence of events during the S2 → S3 transition in photosystem II and implications for the water oxidation mechanism. Proc Natl Acad Sci USA 117(23):12624–12635

Jablonska KM, Ongari D, Moosawi SM, Smit B (2021) Using collective knowledge to assign oxidation states of metal cations in metal-organic frameworks. Nat Chem 13(8):771–777

Kato K, Miyazaki N, Hamaguchi T, Nakajima Y, Akita F, Yonekura K, Shen J-R (2021) High-resolution cryo-EM structure of photosystem II reveals damage from high-dose electron beams. Commun Biol 4(1):382

Kaur D, Szejgis W, Mao J, Amin M, Reiss KM, Askerka M, Cai X, Ibrahim M, Fransson T, Chatterjee R, Cheah MH, Hussein R, Lassalle Nocera DG (2017) Solar fuels and solar chemicals industry. Acc Chem Res 50(3):616–619

Orio M, Pantazis DA (2021) Successes, challenges, and opportunities for quantum chemistry in understanding metalloenzymes for solar fuels research. Chem Commun 57(33):3952–3974

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: machine learning in python. J Mach Learn Res 12:2825–2830

Reeves MG, Wood PA, Parsons S (2019) Automated oxidation-state assignment for metal sites in coordination complexes in the Cambridge Structural Database. Acta Crystallogr B Struct Sci Cryst Eng Mater 75(Pt 6):1096–1105

Retegan M, Krewald V, Mamedov F, Neese F, Lubitz W, Cox N, Pantazis DA (2016) A five-coordinate Mn(IV) intermediate in biological water oxidation: spectroscopic signature and a pivot mechanism for water binding. Chem Sci 7:72–84

Riggs PJ, Yocum CF, Penner-Hahn JE, Mei R (1992) Reduced derivatives of the manganese cluster in the photosynthetic oxygen-evolving complex. J Am Chem Soc 114(26):10650–10651

Shields GP, Raithby PR, Allen FH, Motherwell WD (2000) The assignment and validation of metal oxidation states in the Cambridge Structural Database. Acta Crystallogr B 56(Pt 3):455–465

Suga M, Akita F, Hirata K, Ueno G, Murakami H, Nakajima Y, Shimizu T, Yamashita K, Yamamoto M, Agi H, Shen JR (2015) Native structure of photosystem II at 1.9 Å resolution viewed by femtosecond X-ray pulses. Nature 517(7532):99–103

Suga M, Akita F, Sugahara M, Kubo M, Nakajima Y, Nakane T, Yamashita K, Umena Y, Nakabayashi M, Yamane T, Nakano T, Suzuki M, Masuda T, Inoue S, Kimura T, Nomura T, Yonekura S, Yu L-J, Sakamoto T, Motomura T, Chen J-H, Kato Y, Noguchi T, Tono K, Joty T, Kameshima T, Hatsu T, Nango E, Tanaka R, Naitow H, Matsuura Y, Yamashita A, Yamamoto M, Nureki O, Yabashi M, Ishikawa T, Iwata S, Shen J-R (2017) Light-induced structural changes and the site of O=O bond formation in PSII caught by XFEL. Nature 543(7643):131–135

Suga M, Akita F, Yamashita K, Nakajima Y, Ueno G, Li H, Yamane T, Hirata K, Umena Y, Yonekura S, Yu L-J, Murakami H, Nomura T, Kimura T, Kubo M, Baba S, Kumasaka T, Tono K, Yabashi M, Isobe H, Yamaguchi K, Yamamoto M, Agi H, Shen JR (2019) An oxo/oxo mechanism for oxygen-oxygen coupling in PSII revealed by an x-ray free-electron laser. Science 366(6463):334–338

Tanaka A, Fukushima Y, Kamiya N (2017) Two different structures of the oxygen-evolving complex in the same polypeptide frameworks of photosystem II. J Am Chem Soc 139(5):1718–1721

Umena Y, Kawakami K, Shen J-R, Kamiya N (2011a) Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9 Å. Nature 473(7345):55–60

Umena Y, Kawakami K, Shen J-R, Kamiya N (2011b) Structural changes and the site of O=O bond formation in PSII using time-resolved serial femtosecond crystallography. IUCrJ 8(Pt 3):431–443

Koua FH, Umena Y, Kawakami K, Shen JR (2013) Structure of Sr-substituted photosystem II at 2.1 Å resolution and its implications in the mechanism of water oxidation. Proc Natl Acad Sci USA 110(10):3889–3994

Li H, Nakajima Y, Nomura T, Sugahara M, Yonekura S, Chan SK, Nakane T, Yamane T, Umena Y, Suzuki M, Masuda T, Motomura T, Naitow H, Matsuura Y, Kimura T, Tono K, Owada S, Joty Y, Tanaka R, Nango E, Akita F, Kubo M, Iwata S, Shen JR, Suga M (2021) Capturing structural changes of the S1 to S2 transition of photosystem II using time-resolved serial femtosecond crystallography. IUCrJ 8(3):431–443

Luber S, Rivalta I, Umena Y, Kawakami K, Shen JR, Kamiya N, Brudvig GW, Batista VS (2011) 51-state model of the O2-evolving complex of photosystem II. Biochemistry 50(29):6308–6311

Mino H, Asada M (2021) Location of two Mn2+ affinity sites in photosystem II detected by pulsed electron–electron double resonance. Photosynth Res. https://doi.org/10.1007/s11120-021-00885-5

Nakajima Y, Umena Y, Nagao R, Endo K, Kobayashi K, Akita F, Suga M, Wada H, Naguchi T, Shen JR (2018) Thylakoid membrane lipid sulfoquinovosyl-diacylglycerol (SQDG) is required for full functioning of photosystem II in Thermosynechococcus elongatus. J Biol Chem 293(38):14786–14797

Nocera DG (2012) The artificial leaf. Acc Chem Res 45(5):767–776
complexes in crystal structures of dark-adapted photosystem II. Biochemistry 60(45):3374–3384

Yano J, Yachandra V (2014) Mn₄Ca cluster in photosynthesis: where and how water is oxidized to dioxygen. Chem Rev 114(8):4175–4205

Yano J, Pushkar Y, Glatzel P, Lewis A, Sauer K, Messinger J, Bergmann U, Yachandra V (2005) High-resolution Mn EXAFS of the oxygen-evolving complex in photosystem II: structural implications for the Mn₄Ca cluster. J Am Chem Soc 127(43):14974–14975

Young ID, Ibrahim M, Chatterjee R, Gul S, Fuller F, Koroidov S, Brewster AS, Tran R, Alonso-Mori R, Kroll T, Michels-Clark T, Laksmo H, Sierra RG, Stan CA, Hussein R, Zhang M, Douthit L, Kubin M, de Lichtenberg C, Long Vo P, Nilsson H, Cheah MH, Shevela D, Saracini C, Bean MA, Seuffert I, Sokaras D, Weng TC, Pastor E, Weninger C, Fransson T, Lassalle L, Brauer P, Aller P, Docker PT, Andi B, Orville AM, Glownia JM, Nelson S, Sikorski M, Zhu D, Hunter MS, Lane TJ, Aquila A, Koglin JE, Robinson J, Liang M, Boutet S, Lyubimov AY, Uervirojnangkoorn M, Moriarty NW, Liebschner D, Afonine PV, Waterman DG, Evans G, Wernet P, Dobek H, Weis WI, Brunger AT, Zwart PH, Adams PD, Zouni A, Messinger J, Bergmann U, Sauter NK, Kern J, Yachandra VK, Yano J (2016) Structure of photosystem II and substrate binding at room temperature. Nature 540(7633):453–457

Zhang JZ, Reisner E (2020) Advancing photosystem II photoelectrochemistry for semi-artificial photosynthesis. Nat Rev Chem 4(1):6–21

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.