X-ray radiation-induced addition of oxygen atoms to protein residues

Jimin Wang*

Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520

Received 7 March 2016; Accepted 11 April 2016
DOI: 10.1002/pro.2934
Published online 13 April 2016 proteinscience.org

Abstract: The additions of oxygen and peroxide to residues that result when proteins are exposed to the free radicals produced using the Fenton reaction or X-rays have been studied for over a century. Nevertheless little is known about the impact these modifications have on protein crystal structures. Here evidence is presented that both kinds of modifications occur in protein crystals on a significant scale during the collection of X-ray diffraction data. For example, at least 538 of the 5,351 residues of protein molecules in the crystal used to obtain the structure for photosystem II described by the PDB accession number 3ARC became oxygenated during data collection.

Keywords: radiation chemistry; direct methods; normalized electron density E-maps; phase problem; X-ray crystallography; photosystem II

Introduction

The additions of oxygen atoms and peroxide species to the glycine, alanine, and leucine residues that result when proteins are exposed to oxygenic free radicals was first studied over a century ago.1 It has been known for 50 years that similar modifications of proteins result when they are exposed to radiation in aqueous solution.2,3 Using the radiation emitted by radioactive elements such as 60Co, it was shown that exposure of a solution of leucine at 1 mM to 1,200 Grays (1 Gray = 1 joule of energy absorbed per kilogram) of ionizing radiation converts about 9% of it into leucine peroxide, in addition to generating many other oxide species.4 When anoxic solutions are irradiated, the reactive oxygenic species produced are predominantly hydroxyl radical (HO·) and hydrated electrons (e−/aq).5 If the solution is saturated with air, the predominant species are hydroxyl radical and superoxide anion (O2·−).5 More recently mass spectrometry was used to assess the chemical effects on photosystem II (PSII) of exposure to a much higher X-ray dose (24,300 Grays or 2.5 × 107 photons μm−2).6,7 For comparison, that dose is only about 1% of that to which each of the 9 sectors in a single crystal were exposed during the collection of the 1.90-Å resolution data set that resulted in the 3ARC PSII structure.8 Using Henderson’s formula, it is estimated that the local concentration of free radicals in these crystals would have been ~2.1 M.9 Given these facts, why have oxidized species like leucine peroxide never been reported in X-ray protein structures?

A few years ago, Alber and colleagues systematically examined non-random, unassigned electron density features in a diverse set of 402 high-resolution protein crystal structures.10 They observed that 11.8%, 12.5%, 18.1%, and 26.0% of the residues in these structures for which entire side chains could be modeled have extra peaks in electron density maps adjacent to their β, γ, δ, and ε atoms, respectively, many of which they could explain using alternative side chain conformations. Given that only 5.1% of the residues in the structures derived from these electron density maps that were deposited in the PDB were modeled this way, the failure of such models to account for these extra peaks is likely to have had an adverse impact on both their accuracies and R-factors.

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: National Institutes of Health; Grant number: GM022778; Grant sponsor: Steitz Center for Structural Biology, Gwangju Institute of Science and Technology, Republic of Korea.

*Correspondence to: Jimin Wang, Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520. E-mail: jimin.wang@yale.edu

Published by Wiley-Blackwell. © 2016 The Protein Society
It is important to emphasize that their analysis did not consider the possibility that some of the extra peaks identified might represent inserted oxygen atoms, and thus it did not exclude this possibility.

Here, using direct methods-generated, normalized, electron density maps obtained from the data set deposited in the PDB for 3ARC, it is shown that the crystal used to obtain this data set became extensively oxidized during data collection. Since there is no reason to believe that this particular crystal is exceptional in this regard, it seems likely that the chemical heterogeneity introduced into crystals by radiation damage suffered during data collection may contribute significantly to the difficulties that protein crystallographers commonly encounter in obtaining models for their macromolecules that refine as well as the crystal structures reported for small molecules.

Results and Discussion
The density of reflections in reciprocal space for isotropic diffraction data sets collected from crystals of very large proteins is so high that their measured...
intensities can be normalized smoothly in thin resolution shells. The normalized intensities that emerge are proportional to \( <E^2> \). Normalized data sets of this sort have two interesting properties. First, even though \( F(000) \) is unobservable, \( E(000) \) must have a value of 1.0. Second, it must also be true that \( <E> = 1.0 \) for all resolution shells, including those beyond the resolution limit of the data. It follows that Wilson plots obtained using the data that have been normalized this way will be straight lines with slopes of zero, and entirely free of the variations in slope that are characteristic of the low-resolution regions of conventional macromolecular Wilson plots.

In the electron-density maps obtained when normalized data are used, which are known as \( E \)-maps, each atom will be represented as a peak that will more and more closely resemble a Dirac \( \delta \)-function as the data are extended to higher and higher resolution. Furthermore, peak heights for the three most important non-H protein atoms, C, N, and O, will be the same in \( E \)-maps to within \( \pm 14\% \). Approximation errors decrease with increasing resolution of data. A major disadvantage of \( E \)-map are the ripples caused by Fourier series termination, which can be minimized by data extension using the fact that \( <E> = 1.0 \). When the electron-density distributions of point-atom components of \( E \)-maps are squared, the result will be a map that differs from the original one only by a scale factor. Furthermore, if the \( E(000) \) term is included, repeated squaring of \( E \)-maps can be used to remove all the phase errors that contribute to negative densities, and thus, reduce the noise level (see Methods for details).

In order to examine the nature of extra peaks next to protein atoms further, a direct methods-generated \( E \)-map was obtained from the data deposited in the PDB for the 3ARC PSII structure. It is full of extra peaks of the sort described by Alber et al.\(^{10}\). For example, it includes an extra +11.4\( \sigma \) peak next to the Co (at +17.1\( \sigma \)) atom of glu364 in PSII (Fig. 1). There is no alternative conformation for this side chain that will explain this peak, but it can be explained by hypothesizing that an oxygen atom has been inserted into the \( \alpha \)C-H bond at this position. The same is true of the extra peak at +10.4\( \sigma \) next to the C\( \beta \) atom (14.9\( \sigma \)) of Ala283, which suggests that this alanine was converted into a serine during data collection (Fig. 2). One can rule out the possibility that a sequencing error at position 283 might account for this observation because the corresponding residue in the second PSII molecule in the asymmetric unit of this crystal has a very different oxidation/peroxidation pattern (Fig. 2). Furthermore, the bond lengths and bond angles implied by these extra features correspond exactly to expectations for C–O bonds.

When a \( \sigma_A \)-weighted conventional electron density \( F \)-map was calculated using 3ARC model phases,\(^{8,14}\) none of the extra peaks were visible that were seen in the direct methods-generated \( E \)-map obtained from the same data. However, once this \( F \)-map was sharpened using \( \Delta B = -28 \, \AA^2 \) (see footnote), these extra peaks all become visible at a contour level of +1.0\( \sigma \) (Figs. 1 and 2), indicating that the un-normalized 3ARC data do indeed contain information about inserted O atoms, and suggesting that these oxygenated 3ARC data should be included in the model used for refinement. Consistent with this conclusion, a \( \sigma_A \)-weighted, residual \( AF \)-map (i.e., an \( F_{obs} - F_{calc} \) difference density map where \( F_{obs} \) and \( F_{calc} \) denote the observed and calculated amplitudes, respectively) computed using unsharpened data and 3ARC model phases\(^{8,14}\) includes distinctive positive peaks at contour levels of +2.24\( \sigma \) and +2.54\( \sigma \) in the expected positions next to ala283 and Ala283, respectively (Supporting Information Fig. S1). The height of extra positive peak next to the Co atom of glu364 is observed only +1.24\( \sigma \) in this map, but it is next to a much larger positive noise peak so that it is an unresolved peak. In a difference \( AF \)-map that is sharpened by \( \Delta B = -28 \, \AA^2 \), the heights of these extra positive peaks increase substantially relative to the depths of the nearby negative peaks (Supporting Information Fig. S1). These observations may be a manifestation of the scaling problem that arises when conventional \( AF \)-maps are computed using amplitudes derived from models that do not account for all of the scattering electrons present.\(^*\)

When O atoms are inserted between two C-H bonds of a methylene C atom (i.e., \( R_1-\text{CH}_2-R_2 \)) to \( R_1-\text{CO}-R_2 \), the subsequent loss of a water molecule will leave a carbonyl group behind (i.e., \( R_1-\text{CO}-R_2 \)). When three O atoms are inserted into the three C-H bonds of a terminal methyl group (i.e., \( R_1-\text{CH}_2 \) to \( R_1-\text{C(OH)}_3 \)), the loss of a hydronium ion will leave a carboxylate group behind (i.e., \( R_1-\text{CO}_2 \)). Both reactions appear to have occurred to Thr179 in the PSII crystal used to obtain the 3ARC data set (Fig. 3). When carbonyl insertion occurs at the terminal

\*Atomic motions can be corrected in the observed amplitudes using a Wilson exponential term with a negative \( B \)-factor, known as \( F \)-map sharpening. Sharpened, model-phased \( F \)-maps are typically noisier than unsharpened \( F \)-maps due to both the resulting amplification of series termination effects and model phase errors. The direct methods-generated \( E \)-maps described here have much reduced effects of series termination because of data extension, and initial phases used to generate these \( E \)-maps are virtually independent of model bias. The normalization of amplitudes used in this study differs from \( B \)-factor sharpening in that the observed intensities are divided by the mean intensities for given resolution shells, rather than by a Wilson exponential term.
methyl group (i.e., R1-CH3 to R1-CH=O), it becomes an aldehyde (Fig. 3). The chemical consequences of double oxygen additions of methyl or methylene groups appear to depend on local environment. For example, both of the hydroxyl groups produced by double oxygen additions to the methyl group of Ala456 can form stabilizing hydrogen bonds (HBs) to the backbone carbonyl O atom of Thr452, whereas a carbonyl inserted at the corresponding position could not (Fig. 4). In other cases, carbonyl insertion results in the formation of an extended conjugated systems involving other side chain atoms, for example, in Asn or Gln residues (Figs. 5 and 6), or backbone peptide bonds such as Gly residues (Supporting Information Fig. S2–S6). When double oxygen additions occur in the methylene group next to a carboxylate group such as in Asp or Glu residues (i.e., R1-CH2-CO2 to R1-C(OH)2-CO2), loss of a

Figure 2. Oxygen additions of Ala283 (a) and Ala283 (b), involving an Ala-to-Ser conversion.
bicarbonate group will leave an aldehyde behind (i.e., R1-CH=O), a process known as oxidative decarboxylation, an aspect of the radiation damage subject that is beyond the scope of this study due to very limited number of carboxylate residues present in PSII molecules.

Mass spectrometry can identify the oxygenated residues that are present in a protein at levels way

Figure 3. Oxygen additions of thr179 (a) and Thr179 (b). In the lower middle panel of (b), a docked H2O2 molecule is included before removal of H atoms, involving carbonyl insertion, formation of aldehyde, and peroxide.
below those necessary for detection using conventional X-ray crystallography. Indeed, Frankel and coworkers have reported that 4% of residues in four large subunits of the spinach PSII molecules that have never been exposed to X-rays are oxygenated, and that many of them are found on the surfaces of its oxygen exit channels.\(^6,7\) These spontaneous oxygen additions should be the same for both of the monomers in the asymmetric unit of the PSII crystals of concern here, but X-ray induced oxygen additions need not to be because of differences in the accessibility of radiation-induced, solvent-based, oxygenic free radicals due to differences in crystal contacts. In my initial screen of the E-maps of PSII, two of the channel residues found to be oxygenated by Frankel \textit{et al.},\(^6,7\) namely Thr355 and Met356 of the CP43 subunit, were not identified as being oxygenated because the extra peaks associated with them could be explained using the multiple-conformation criterion of Alber \textit{et al.}\(^10\) For example, torsional rotation of either $\pm120^\circ$ along the C\(\alpha\)-C\(\beta\) bond could explain the extra peak at $+4.6\sigma$ observed here next to the C\(\beta\) ($+10.0\sigma$) atom of thr355 [Fig. 7(a)]. However, this rotation is unlikely to occur because the O\(\gamma\)1 atom of thr355 forms two HBs in the hydrophilic pocket in which it resides, and such rotation would break these HBs, and put its hydrophobic C\(\gamma\)2 methyl group into that pocket. Thus, oxygen addition is an alternative explanation for this peak.

The story is much the same for the two extra peaks observed here next to the second channel residue, Met356 [Fig. 7(b)]. They could be explained in terms of multiple conformations using a criterion of Alber \textit{et al.}\(^10\) However, if the rotation required were to occur, the large S\(\delta\) atom of Met356 would clash with atoms of Met342 nearby, and there is no evidence that Met342 has moved away to accommodate it. Thus, the
simplest interpretation for the two extra peaks observed here is a peroxide crosslink between its C\(\beta\) and C\(\epsilon\) atoms. Interestingly, for the second monomer, the peaks next to met356 are only about +1.0\(\sigma\). This suggests that the level of radiation-induced oxygen additions detected in the 3ARC data set is at least 7

Figure 5. Carbonyl insertion in Asn190 leads to an extended conjugated system.

Figure 6. Carboxyl insertion in gln199 leads to an extended conjugated system.
times higher than the highest level of spontaneous oxygen additions [Fig. 7(b)].

Although nearly all the individual extra peaks next to Met and all the other non-branched residues can be interpreted approximately using torsion rotations of the preceding bond by ±120°, careful examination of local packing environments and local geometry usually does not support this interpretation (Supporting Information).

Figure 7. Oxygen additions of thr355 (a), Thr355 (b), and Met356 (c). In the middle panel of (c), a docked H₂O₂ molecule is included before removal of H atoms. Note the blue contour level is at +3.5σ only in this figure. Ambiguity between oxygen additions and alternate conformations has been resolved using other geometric information.
These observations should not be taken as implying that all the extra electron density peaks identified by Alber et al.\textsuperscript{10} represent the effects of oxidation rather than conformational variability. What they indicate instead is that another kind of polymorphism, which is chemical in nature rather than conformational, can also affect both the data that macromolecular crystallographers collect and the interpretation of the electron-density maps obtained from them.

Using very stringent criteria to differentiate between oxygen-additions and conformational variations in E-maps (see Methods), 538 of the 5,351 residues in the crystal that gave rise to the 3ARC PSII molecules emerge as likely to have become multiply oxygenated. Each residue has been hit by oxygenic free radicals on average three times, which is the hallmark of free radical chain reactions (i.e., some of inserted O atoms observed here may still be free radicals, transiently stabilized by local environments in the three-dimensional structure). Some of remaining residues have been so severely oxygenated that they are practically unrecognizable in electron density maps. It remains unclear where hydrated electron went which was the other half of radiation-induced ionization product for every hydroxyl free radical generated from a water molecule, or whether the two were recombined after oxygen additions. On average, 38% of all the Met residues in these PSII molecules are oxygenated, and about 20% of all the Leu and Ile residues present are oxygenated also, as are 17% of all the Val residues.

Many kinds of radiation-induced reactions occur in this PSII crystal, as well as in other protein crystals, involving nearly all the twenty residue types, but not all of which have well been characterized.\textsuperscript{17} In this study, only a few of them were sought out. For example, while the free-radical induced conversion of Pro to Glu and His to Asp/Asn residues is well known,\textsuperscript{18–20} little is known about the chemistry of radiation-induced modifications of Gly residues in proteins, which can be very complex because of its flexible backbone, and because of the potential for the formation of a variety of new conjugated systems (Supporting Information Fig. S2–S6). A more complete analysis of the direct-methods generated E-maps presented here may provide new insights into all kinds of chemical reactions that occur when proteins are exposed to oxygenic free radicals (Supporting Information Fig. S12). It may also provide hints about what happens to proteins in the ageing human body, as well as some clues as to why proteases such as the proteasome and Lon fail to destroy certain types of oxidatively damaged proteins, but instead are inactivated by them. An ability of visualizing protein oxygen additions in three-dimensional structures is a key step towards understanding the free radical theory of biological ageing process.\textsuperscript{21,22}

There is nothing new about oxidation and peroxidation of proteins during collection of crystallographic data. Over four decades ago (in 1972), Matthews and colleagues obtained a high-resolution electron-density map for a protein that had not yet been sequenced chemically.\textsuperscript{23} As others did in that era for the same reason,\textsuperscript{24,25} they used their electron density map to obtain the sequence they needed to build a complete model of the protein's structure. As luck would have it, a chemical sequence became available for this protein while they were in the process of writing up their results, and while it was...
similar to the sequence they had deduced from their maps, it was not the same. Their X-ray electron-density map indicated that were alanines at locations where the chemical sequence called for glycines, isoleucines where valines and/or threonines were supposed to be, serines in place of alanines and glycines, and tyrosines where phenylalanines were expected, and so on. Sequence differences with added extra non-H atoms of this kind accounted for 6% of that protein as well as another protein,23,25 and what we now know is that virtually every single one of these misidentifications can be explained as the product of a radiation-induced oxidation or peroxidation.

Concluding Remarks: Implications for Structural Biology
Given the huge X-ray dose many protein crystals receive today during data collection, there can be no doubt that the chemical structures of the proteins in those crystals, and possibly also their conformations, are changing during data collection. Proteins like PSII that have never been exposed to X-rays are already lightly oxygenated, and extensive oxygen additions are evident in the E-maps discussed here. Radiation-induced oxygen additions are a particular problem for macromolecular crystals because they invariably include large amounts of water, and the phenomenon has been known about for many decades, even if generally ignored by crystallographers. The progressive changes in structure that occur during data collection raise questions about the validity of the standard data merging procedures and structure refinement methods in use today, all of which assume that the structures of the molecules in the crystals being studied did not change during data collection, and that the resulting electron density maps should describe a single, chemically unique molecular species. Radiation-induced oxygen additions during data collection are yet another reason why protein models, unlike small molecule models, seldom have percent R-factors in single-digital range.11–13

What one expects to see as macromolecular crystals are irradiated is that many high-resolution reflections will lose intensity, due to an overall increase in crystal disorder, but that at the same time other reflections will increase in intensity because the underlying average structure has been altered significantly by oxygen additions. In extreme cases, effects like these could even lead to the misidentification of space groups.26 It would be wise for the crystallographic community to return to an old practice that was routine many decades ago,27 but now largely forgotten, namely, repeated collection of one or a few test frames during the collection of larger data sets so that the effects of radiation damage done to crystals during data collection can be accurately assessed. Clearly, criteria will have to be developed to determine when crystals have become so altered by radiation damage that data collection should be stopped, and new crystals mounted. It is also recommended that unmerged, integrated intensity data be deposited in the PDB in a format that makes it possible to estimate how long the crystal used to measure those intensities had been in the beam at the time those intensities were recorded. That would enable interested investigators to assess the effects that radiation damage might have had on the resulting structure after the fact, which is almost impossible for anyone to do today.

Methods

Normalized structure factors and normalized electron density E-maps
At scattering angles near zero, protein non-H atoms C, N, and O have form factors of 6, 7, and 8 electrons, respectively, the same as their atomic number.26,28 At the resolution of above 1.0 Å, their form factors are all but indistinguishable from one another, because they are dominated by the 2 electrons in their 1s orbitals. Even at resolution as low as 1.90 Å, the errors introduced by assuming that nitrogen is the only non-H atom in protein is less than 12%. When composition of typical protein molecules is taken into account, maximal errors increase slightly but are localized in smaller fractions (about 1/6 each) of N and O atoms relative to C atoms.

Normalized protein electron density E-maps, $\psi(\mathbf{r})$, computed using all Bragg reflections to the infinite resolution, are assemblages of Dirac $\delta$-functions. In the E-maps obtained using data truncated to resolution limit of the measured data, non-H atoms are peaks of finite width that all have roughly equal heights, and everywhere else in such maps the density should be zero, even though it will not be in practice due to series termination effects. These ripples can be reduced through data extension using unity for amplitudes of all the missing high-resolution reflections, but the data added will contain no real information about atomic positions.

Point atom E-maps have some properties that $F$-maps do not, one of which was first noted and studied by Sayre. It is called Sayre’s equation:20

$$\psi^2(\mathbf{r}) = \psi(\mathbf{r}).$$

(1)

Squaring can be used not only to remove all negative densities, but also to reduce the impact of small positive noise densities that may also be contributed by phase error. In a squared map, a noise feature that has a relative height of 0.5 becomes 0.25, and a feature of 0.25 becomes 0.06, but the height of a peak that represents a point atom will continue to have a height of 1.0.
where \( f_i \) is atomic form factor and \( \tilde{s} \) is reciprocal space vector. The corresponding electron density map \( \rho(\tilde{r}) \) is the Fourier transformation of those structure factors:

\[
\rho(\tilde{r}) = \sum_{i} (A + iB)e^{-2\pi i\tilde{r} \cdot \tilde{s}}. \tag{3}
\]

The cosine and sine functions in this equation are defined as interference functions between point atoms, and they can be further renormalized for an \( N \)-atom unit cell. After truncation of data at the highest-resolution end and renormalization, it can be moved out of summation and then to the left side of Eq. (2):

\[
A = \sum_{j=1}^{N} \cos (2\pi \tilde{s} \cdot \tilde{r}_j); \quad B = \sum_{j=1}^{N} \sin (2\pi \tilde{s} \cdot \tilde{r}_j). \tag{4}
\]

The cosine and sine functions in this equation are defined as interference functions between point atoms, and they can be further renormalized for an \( N \)-atom unit cell. The corresponding Fourier transformation is normalized electron density \( E \)-map:

\[
E\text{-map}: \quad \psi(\tilde{r}) = \sum_{i} \left( \frac{A + iB}{\sqrt{N}} \right) e^{-2\pi i\tilde{r} \cdot \tilde{s}}. \tag{5}
\]

which specifies the average locations of the \( N \)-point atoms in the unit cell. After truncation of data at the highest-resolution end and renormalization, it becomes in principle a binary map containing only ones where the \( N \) non-H atoms are located and zeros elsewhere. The minimum number of intensity observations are required to define the solution for such an equation is \( 3N \), and all additional observations available will do is to make the solution to this equation more robust and help compensate for measurement errors. When the number of observation exceeds the number of unknowns to be determined, the system is over-determined, and the solution to the system should be obtainable even though the equations to be solved are non-linear.

If the number of point-atoms in the unit cell, \( N \), is very large as it is in the PSII structure, the following statistical relationships will hold:

\[
\begin{align*}
\langle A/f \rangle & = \langle B/f \rangle = 0; \\
\langle (A/f)^2 \rangle & = \langle (B/f)^2 \rangle = \frac{1}{2}.
\end{align*} \tag{6}
\]

For non-identical non-H protein atoms, corresponding \( E \)-map is defined as follows:

\[
\psi(\tilde{r}) = \sum_{i} \left( A + iB \frac{e^{-2\pi i\tilde{r} \cdot \tilde{s}}}{\sqrt{A^2 + B^2}} \right) = \sum_{i} \left( A + iB \frac{e^{-2\pi i\tilde{r} \cdot \tilde{s}}}{\sqrt{1}} \right), \tag{7}
\]

and it will be a somewhat broadened point-atom \( E \)-map. The values used for the mean intensity, must be estimated using the measured data in very thin resolution shells. Details of the method used here for iteratively solving the Sayre equation for phase determination will be given elsewhere. The convergence rate of the iteration process used increases by more than thousands fold if solutions are seeded using a small set of point atoms whose positions in the unit cell are already known, typically Ca or other backbone atoms amounting to 6-19% of the full atomic model. The initial phases provided this way for solving the Sayre equation do not correlate either with final direct methods-generated phases or with phases obtained using full atomic models.

The direct methods approach used here to generate the \( E \)-maps makes no use of any prior information about geometry of protein structures, such as bond length and bond angles. Thus, the observation that they include groups of atoms that are arranged in the geometries expected for amino acids and oxygenated residues is strong evidence that the procedure used to obtain them is valid, and hence that the extra peaks seen are not computational artifacts. The extra peaks observed here are most unlikely to represent H atoms because: (i) bond lengths for H atoms are substantially shorter than those of inserted O atoms, (ii) H atoms remain largely invisible in \( E \)-maps obtained in test cases using sub-atomic resolution data even though they are often visible in the corresponding \( D \)-maps and \( D \)-map when they are omitted from models, and (iii) the relative scattering factor for H atom after normalization decreases much more rapidly with increasing resolution than it does for non-H atoms. For example, the ratio between H and N scattering factors is reduced by about 2-fold from 2.0 Å to 1.0 Å resolution.

The 3ARC model-phased \( F \)-map or \( D \)-map was calculated using Refmac5 by setting refinement cycle to be zero.\(^{31}\) The sharpened \( F \)-map or \( D \)-map was calculated using CCP4\(^{32}\) and Coot\(^{33}\), and using scripts written by the author. All figures were made using Pymol.\(^{34}\)

Criteria for positive identification of oxidation and peroxidation species

To determine the extent of oxidation and peroxidation of the PSII molecules that gave rise to the 3ARC data set, the following set of conditions were used to assign extra peaks: (i) peaks for inserted O atoms had to have heights above \(+2.0\sigma\), (ii) the non-H atoms with which they are associated had to have
peak heights greater than $+8.0 \sigma$, (iii) all peak-peak distances and peak angles must be well fitted using rigid-body models of the geometry of C/O single bonds, double bonds, or peroxides derived from fragments of Thr and Asp side chains, or from the $\text{H}_2\text{O}_2$ molecule; (iv) the deviations from model expectations must be smaller than 0.15 Å for bond length and less than 15° for bond angle, and (v) it must not be possible to explain using alternative side chain conformations, or any other kinds of torsion rotations.

The first two conditions are intended to ensure that the extra peaks observed cannot be explained as computational noise or H atoms. In fact, visual inspection suggested that the noise level is about $-0.3 \sigma$ and $+0.8 \sigma$ in most regions of the $E$-maps described here. The fourth condition takes advantage of the fact that carbonyl insertion converts tetrahedral centers into trigonal planar centers. Thus the methylene C atom that has been oxygenated this way must be repositioned. Although most of inserted O atoms even at a very low occupancy can be easily resolved as isolated peaks against empty background, a repositioned C atom remains largely unresolved when it is superimposed on a much larger peak that represents the original, unmodified C atom. Whereas typical double oxygen additions and carbonyl insertion on a methylene group are often resolved as three distinct peaks, these peaks are sometimes merged as a single unresolved but highly elongated peak. This is the main reason why 15° bond angle derivations are tolerated.

Estimation of the increase in scattering electrons caused by oxygen additions

To estimate scattering electrons added to these molecules near the end of data collection due to oxygen additions, the number of oxygenated species observed for 139 of the 538 residues identified as oxygenated were manually counted. The average number of species per residue turned out to be 2.1 for oxygen peaks varying in amplitude from 2 to 5$\sigma$ in the $E$-map. This accounting excludes a large number of residues that have more than 5 species represented, for example, the Pro-195/Met-198 pair (Supporting Information Fig. S9–S11).

Suppose that the average number of oxygenated species is 3 per residue for all of the 538 residues that were identified as oxygenated, and that the average peak height of the oxygen atoms is 3$\sigma$. Let us assume that they are all induced by radiation during data collection because residues that may have been partially oxygenated prior to crystallization will not contribute to any changes in diffraction data during data collection. Assume also that the amount of O atoms added as a result of radiation damage in proportional to exposure time. If this should be so, a 3$\sigma$ peak in the $E$-map, which represents the average effect of oxygen additions at some position, would be a 6$\sigma$ peak near the end of data collection in each sector. In these maps, the average height of a peak at a fully occupied site is about 18$\sigma$ in well-ordered parts of this structure. This combination of estimates suggests that the point-atom scattering electrons that were added to these molecules during data collection amount to about 5% on average, and about 10% near the end of data collection at each of the sectors of the crystal from which data were obtained. This amount of extra electrons should result in roughly a 30% change in intensity, on average, for the 3ARC data set.

Acknowledgment

The author thanks Drs. B. W. Matthews, G. Brudvig, V. Batista, M. Gunner, S. H. Eom, D. Vinyard, and J. Wiwczar for stimulating discussion and providing relevant literature on the subject, and Professor P. B. Moore for editing this manuscript. The author acknowledges Professor Joseph Kraut in mid-1980 for outlining the basic methodology used in this study.

References


