Molecular oxygen binding in the mitochondrial electron transfer flavoprotein

Peter Husen,† Claus Nielsen,† Carlos F. Martino,‡ and Ilia A. Solov’yov*,¶

†Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, Odense, Denmark
‡Biomedical and Chemical Engineering and Science Department, Florida Institute of Technology, Melbourne, Florida
¶Department of Physics, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany

E-mail: ilia.solovyov@uni-oldenburg.de
Abstract

Reactive oxygen species such as superoxide are potentially harmful byproducts of the aerobic metabolism in the inner mitochondrial membrane, and complexes I (the ETF enzyme) and III (the cytochrome bc1 complex) of the electron transport chain have been identified as primary sources. The exact mechanisms of superoxide production have not been fully established, but a crucial starting point would be the binding of molecular oxygen within one of the protein complexes. The present investigation offers a comprehensive computational approach for the determination of binding modes and characteristic binding times of small molecules inside proteins, which is then used to reveal several O2 binding sites near the flavin adenine dinucleotide cofactor of the ETF enzyme. The binding sites are further characterized to extract the necessary parameters for further studies of possible electron transfer between flavin and O2 leading to radical pair formation and possible superoxide production.
Introduction

Every aspect of life on the planet involves electron transfer flavoproteins (ETFs). These proteins are found mostly in mitochondria and are not well characterized due to their complexity, but appear to be the common sites of accepting electrons and interacting with them. ETFs are known to be heterodimeric proteins that contain two subunits: a larger alpha and a smaller beta subunit. At least a dozen mitochondrial Acyl-CoA dehydrogenases containing flavin adenine dinucleotide (FAD) are in contact with respiratory electron transfer through ETF-ubiquinone oxidoreductase (ETF:QO), which is located in the mitochondrial matrix together with the ubiquinone pool and thus directly influence the generation of ATP.

Under normal conditions, reactive oxygen species (ROS) such as superoxide are generated in the inner mitochondrial membrane as a byproduct of aerobic metabolism with complexes I and III from the electron transport chain (ETC) as the major sources of superoxide production. ROS generation by these complexes depends strongly on the redox rate in the ETC. Redox chemistry, however, not only happens in the main respiratory complexes, but also in other places, such as ETF and ETF:QO as indicated in Fig. 1. These proteins thus appear to be likely additional sources of ROS production. In particular, superoxide is thought to be formed in ETF:QO with the exact mechanism being uncertain, but likely associated with either the ETF:QO pool system or semiquinone forms of FAD. It is hypothesized that superoxide produced inside mitochondria is unable to escape through the mitochondrial membrane and may damage the structure of the mitochondria unless converted to hydrogen peroxide.

In mammalian cells, reactive oxygen species production stems from four enzymatic systems: 1) NADPH oxidases (Nox); 2) xanthine oxidase; 3) uncoupled NO synthase, and; 4) the mitochondrial electron transport chain complexes. Many biological processes,
Figure 1: Electron transfer from octanoyl-CoA to ubiquinone involves flavoenzymes medium-chain acyl-CoA dehydrogenase (MCAD), electron transfer flavoprotein (ETF), and ETF-ubiquinone oxidoreductase (ETF:QO). ROS may be formed as a byproduct of electron transport in fatty acid oxidations.ROS may be formed as a byproduct of electron transport in fatty acid oxidations.

such as proliferation and bioenergetics, are regulated by the crosstalk between these ROS generating centers, where ROS products, such as hydrogen peroxide and superoxide, act as signaling molecules. Superoxide is often a precursor for other ROS species such as hydrogen peroxide (H$_2$O$_2$), peroxynitrite (ONOO$^-$), lipid and hydroxyl radicals ($\cdot$OH), where overproduction of O$_2$$^\bullet$ is involved in oxidative stress. However, a new paradigm is beginning to emerge that under normal physiological conditions O$_2$$^\bullet$ and H$_2$O$_2$ are important signaling molecules that control specific biochemical reactions and metabolic pathways.

Oxidized flavins are formed after the autocatalytic reaction of reduced flavins with solubilized oxygen. The initial formation of a complex between the reduced flavin and oxygen is required in order to have consistent kinetic behavior. This complex has been identified as C(4a)-flavin hydroperoxide. The reaction starts with the formation of a caged radical pair (semi-reduced flavin–superoxide anion) by a single electron transfer from the reduced flavin to the molecular oxygen. This reaction proceeds close to C4(a). The radical pair can then either 1) form hydrogen peroxide, 2) insert an oxygen atom into another molecule or 3) produce oxygen radicals (i.e. superoxide anion). Considerable evidence suggests that cellular bioenergetics may have important causal effects in disease susceptibility including...
aging,\textsuperscript{16,17} inflammation,\textsuperscript{18} diabetes and obesity,\textsuperscript{19,20} neurodegeneration,\textsuperscript{21,22} infection,\textsuperscript{23} and cancer.\textsuperscript{24,25} For example, mitochondrial dysfunction is now recognized as a key element in neurodegenerative diseases including Alzheimer’s\textsuperscript{26} and Parkinson’s disease.\textsuperscript{22,27} Reactive oxygen species overproduction is also implicated in various diseases\textsuperscript{28} including neurological damage,\textsuperscript{29} viral replication,\textsuperscript{30,31} immune dysfunction,\textsuperscript{18,32} and programmed cell death.\textsuperscript{33} Therefore, modulating specific ROS products can play a key role in cellular bioenergetics, mitochondrial function,\textsuperscript{22} and underlying disease processes \textsuperscript{16,29} including Alzheimer’s disease.

The exact sites and mechanisms of ROS production in the mitochondrial protein complexes remain controversial. The starting point for an efficient superoxide production mechanism would be the binding of O$_2$ within a redox-active protein, and the focus of the present study is to investigate possible binding of O$_2$ within the ETF. Although other oxidation states of the FAD cofactor could also be relevant for ROS production in flavoproteins,\textsuperscript{10} the present study focuses specifically on possible ROS production mechanisms starting from the fully reduced state of the cofactor.\textsuperscript{4,15} Computational modeling based on molecular dynamics (MD) of the ETF in presence of molecular oxygen is employed to localize and characterize possible O$_2$ binding modes, which could be relevant for superoxide production. MD simulations have previously been applied to study the interactions of small molecules and ions with proteins and membrane-protein systems,\textsuperscript{6,34,35} and this study establishes a protocol dedicated to identify and characterize binding sites of small molecules within a protein. Binding sites and characteristic binding times of O$_2$ near the FADH$_2$ cofactor are of particular interest, as it is the redox active part of the ETF enzyme and could serve as a donor in an electron transfer leading to superoxide (O$_2^-$) production. Uniquely, hundreds of simulations are carried out to generate sufficient statistics and to allow quantitative characterization of the identified oxygen binding modes. This investigation presents the basic approach that could be easily generalized to identify pos-
Methods

Molecular dynamics simulation protocol

The computational modeling in the present investigation was based on all-atom MD simulations supported by quantum chemical calculations used to obtain force-field parameters for the fully reduced flavin adenine dinucleotide (FADH$_2$) cofactor. Separate sets of MD simulations were dedicated to charting of O$_2$ localization and to the detailed investigation of O$_2$ binding at each of the five identified candidate binding sites. The former employed high oxygen concentrations to achieve good statistics, while the latter included only a single
Table 1: Summary of parameters used for the molecular dynamics simulations.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature (T)</td>
<td>300 K</td>
</tr>
<tr>
<td>Salt concentration (c_{NaCl})</td>
<td>50 mM (NaCl)</td>
</tr>
<tr>
<td>Waterbox size (L_x \times L_y \times L_z)</td>
<td>87.3 Å x 98.3 Å x 106.4 Å (equilibrated)</td>
</tr>
<tr>
<td>Waterbox volume (V)</td>
<td>912 nm^3</td>
</tr>
<tr>
<td>Number of O_2 molecules (N)</td>
<td>10, 20, 30 and 40</td>
</tr>
<tr>
<td>O_2 concentrations (c_{O_2})</td>
<td>18.2 mM, 36.4 mM, 54.6 mM and 72.8 mM</td>
</tr>
<tr>
<td>Number of replicates</td>
<td>10 per O_2 concentration</td>
</tr>
</tbody>
</table>

O_2 molecules in the simulation box. The computational protocols are outlined below.

The system consisting of the ETF protein (PDB ID 1EFV^{36}) with bound FADH_2 and adenosine monophosphate (AMP) was modeled by all-atom MD simulations using the NAMD software package.\textsuperscript{37} The combined system was solvated in a water box extending 18 Å away from the protein on each side, leading to a water box size of approximately 87 Å x 98 Å x 106 Å after equilibration. A number of O_2 molecules were added in the bulk water by random replacement of water molecules.\textsuperscript{6} The parameters used for the simulations are summarized in Table 1, and the simulation box is illustrated in Fig. 2. The CHARMM 36 force field with CMAP corrections\textsuperscript{38-40} was used for modeling the protein, solvent and the AMP cofactor. The FADH_2 cofactor was parametrized by modification of an earlier model of FADH\textsuperscript{•}\textsuperscript{41} as outlined in the following section. Molecular oxygen was modeled following an earlier study.\textsuperscript{6}

The system was equilibrated, first under constant pressure (NPT ensemble) to allow the simulation box to shrink and acquire a physiological density, while further equilibration continued through employing constant volume simulation (NVT ensemble). Constraints on the system were gradually released: first, only the solvent was equilibrated, then the side chains and cofactors of the ETF were released, and finally, the protein backbone was released as well to equilibrate the whole system. After equilibration, four different concentrations of O_2 molecules were added to the bulk water: a total of either 10, 20, 30 or
Table 2: Simulation protocol. The system was equilibrated under constant pressure (NPT ensemble) and then under constant volume (NVT ensemble). Constraints on the system were gradually released. After equilibration, ten replicate simulations of 100 ns each were performed for each of the four O₂ concentration. In the O₂ binding event simulations, 100 replicate simulations were carried out for each of five separate identified putative O₂ binding sites with an O₂ molecule initially placed in the binding site.

<table>
<thead>
<tr>
<th>Statistical ensemble</th>
<th>Integration timestep</th>
<th>Constrained atoms</th>
<th>Repetitions</th>
<th>Simulation time</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Initial equilibration:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPT</td>
<td>1 fs</td>
<td>Protein and cofactors</td>
<td>1</td>
<td>2 ns</td>
</tr>
<tr>
<td>NPT</td>
<td>1 fs</td>
<td>Protein backbone</td>
<td>1</td>
<td>3 ns</td>
</tr>
<tr>
<td>NPT</td>
<td>1 fs</td>
<td>none</td>
<td>1</td>
<td>1 ns</td>
</tr>
<tr>
<td>NPT</td>
<td>2 fs</td>
<td>none</td>
<td>1</td>
<td>220 ns</td>
</tr>
<tr>
<td>NVT</td>
<td>2 fs</td>
<td>none</td>
<td>1</td>
<td>50 ns</td>
</tr>
<tr>
<td><strong>O₂ localization simulations:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPT</td>
<td>2 fs</td>
<td>none</td>
<td>4</td>
<td>5 ns</td>
</tr>
<tr>
<td>NVT</td>
<td>2 fs</td>
<td>none</td>
<td>4 × 10</td>
<td>100 ns</td>
</tr>
<tr>
<td><strong>O₂ binding event simulations:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NVT</td>
<td>2 fs</td>
<td>none</td>
<td>5 × 100</td>
<td>1–25 ns</td>
</tr>
</tbody>
</table>

40 O₂ molecules were added to the simulation box. The corresponding concentrations are provided in Table 1. A short additional equilibration in the NPT ensemble was performed for 5 ns for each O₂ concentration. After equilibration, ten replicate simulations of 100 ns each in the NVT ensemble were performed for each O₂ concentration. Additionally, a number of simulations dedicated to analyzing O₂ binding times were carried out. The complete simulation protocol is outlined in Table 2.

The simulated trajectories were analyzed to identify O₂ localization, binding sites and binding events. The software packages VMD⁴² and GNU Octave⁴³ were used for analysis and visualization.
Parameterization of FADH$_2$

The FADH$_2$ cofactor was parameterized based on an earlier model of FADH$^*$. The flavin adenine dinucleotide was fully reduced by addition of an extra hydrogen. Partial charges, bonded and angular force field parameters were then optimized using the Force Field Toolkit (ffTK) in VMD. Quantum chemical calculations were carried out using Gaussian09, and the results were used to construct force field parameters through the fitting routines of ffTK. Water interaction calculations for fitting partial charges were carried out using the Hartree-Fock method (HF) as implemented in Gaussian09, while geometry optimization and bonded interaction calculations were using second order Møller-Plesset corrections (MP2). All quantum calculations made use of the 6-31G* basis set. The obtained partial charges and force field parameters are provided in the supporting information (SI).

Simulation of binding events

Following the simulations described in the previous section, five potential oxygen binding sites were identified, and a series of 100 additional simulations per binding site were carried out in order to estimate the characteristic binding time. A special simulation protocol was used to assess the binding time at each of the five binding sites, starting with a configuration where an oxygen molecule was bound. All of the additional oxygen molecules were removed from the simulation box for the five bound configurations, such that cooperative binding effects due to the unnatural concentration of oxygen could be eliminated. Removal of the oxygen molecules might slightly perturb the system, so 500 energy minimization steps followed by a 100 ps simulation in the NVT ensemble was performed in order to equilibrate the five bound configurations, assuming that the removal of the small oxygen molecules had a negligible effect on the volume of the system. Thus, one configuration for each binding
site with only the single bound oxygen molecule in the simulation box was obtained. For improved statistical independence, five separate starting configurations were generated for each of the five binding sites by taking five snapshots distributed over an initial 800 fs simulation. Both the coordinates and the velocities of all atoms were saved for each starting configuration, since a re-initialization of the velocity in the following simulation step would lead to incorrect kinetic energies for the bound oxygen molecules and could possibly affect the measured binding times. Each starting configuration was simulated independently 20 times until the oxygen molecule left the binding site, such that a total of 100 independent simulations per binding site, starting with bound oxygen, was performed. Specifically, the simulations were terminated, when O$_2$ moved more than 10 Å away from its initial position, thus providing sufficient statistics for the analysis detailed below.

**Unbinding criterion**

Determination of the binding time requires a clear definition of what it means for an oxygen molecule to be bound inside the ETF enzyme, and will necessarily rely on some assumptions. Here, it is postulated that a binding site is a spherical region defined by a binding site center and a binding radius. In the simulations of binding events at each binding site, the oxygen molecule was initially placed within the binding region and then considered bound as long as it remained inside. Once unbinding occurred, i.e. the O$_2$ molecule left the binding region, the simulation was stopped and the elapsed time until the unbinding event was recorded.

Both the center and the radius for each binding site were determined in a three-step process. The first step was to select an approximate radius that covers the binding volume. In order to do so, a binding radius of 10 Å was assumed with the initial position of the oxygen molecule being the approximate center of the binding site, i.e. the initial position of the oxygen molecule in each of the 5 starting configurations per binding site was used as
the binding center for each of the corresponding 20 independent binding event simulations. The distance from the oxygen molecule to the chosen binding site center was then measured throughout the simulations.

For the second step, the binding site radius was refined based on the calculated distribution of distances between the oxygen molecule and the binding site center obtained in the first step. Shrinking the binding site radius in this way was necessary since the binding distance of 10 Å in the first step was large enough to potentially include multiple binding sites, which would lead to artificially long binding times. The simulated trajectories were truncated according to the refined unbinding criterion, and the average position of the bound O\textsubscript{2} was calculated. This average position was then used as a refined binding site center for the third step, in which the final unbinding times and O\textsubscript{2} positions were sampled.

**Results**

Here, we present the results of the simulations and the associated analysis. First, the observed localization of O\textsubscript{2} within the ETF protein is described. Then, the results of simulations dedicated to characterizing each of five identified candidate O\textsubscript{2} binding sites are presented and used to measure binding times based on a Poisson process model. Finally, the candidate binding sites are described in more detail by taking a closer look at the local binding environment.

**Distance of closest approach**

One way to discover binding sites of oxygen near FADH\textsubscript{2} is to measure the distance of closest approach, i.e. the smallest distance between an atom from any of the oxygen molecules and an atom from the isoalloxazine moiety (flavin) of FADH\textsubscript{2}. The flavin group was used for this measure, as it is expected to be the redox active part of the FADH\textsubscript{2} cofactor.\textsuperscript{2} The
edge-to-edge distance is the de facto measure for electron transfer studies, since the electronic wavefunction is likely to be delocalized over the ring structure of the flavin group, so proximity to any part of flavin could lead to an electron transfer reaction.49,50 The flavin–O$_2$ distance was sampled for each O$_2$ molecule during the simulations with many O$_2$ molecules, which allowed to construct the flavin–O$_2$ distance distribution illustrated in Fig. 3. A number of peaks in the distributions suggest the presence of possible O$_2$ binding sites within the ETF protein. Figs. S1 and S2 in the SI shows the distance of closest approach between flavin and the nearest O$_2$ molecule as a function of simulation time, and here, actual binding events are evident as plateaus in the distance graphs, indicating that an O$_2$ molecule is confined to a small region of space for up to tens of nanoseconds at a time. Fig. 3 also shows the measured distance between flavin and the five binding sites identified in the following analysis. These binding sites are not all clearly resolved in the distance distribution, which is primarily due to the fact that the distributions in Fig. 3 are taken over surface shells at increasing distances from the flavin, while a binding site only occupies a very small part of such a shell.

The modeled concentrations of O$_2$ are several orders of magnitude greater than physiological conditions and were used to obtain sufficient statistical sampling from the limited simulation times realistically obtained by MD simulations. The four different concentrations were then used to gauge possible artifacts due to the high concentration through e.g. cooperative effects. The distance distributions in Fig. 3 for the four different O$_2$ concentrations show many of the same features, but also a noticeable variation, indicating that this first set of simulations does not by itself provide sufficient statistics for a complete quantitative characterization of O$_2$ binding in the ETF protein, but can be used to detect candidate binding sites for further analysis. The peaks in the plots in Fig. 3 for 30 and 40 O$_2$ molecules in the range between 8 and 16Å appear less pronounced than for smaller concentrations, indicating that the ETF protein may be getting saturated with O$_2$. 
Figure 3: The distribution of edge-to-edge distances between any single O$_2$ molecule and the isoalloxazine (flavin) moiety of FADH$_2$ over all replica of the O$_2$ localization simulations, see Table 2. The average distance between flavin and the bound O$_2$ molecule obtained in the separate set of simulations to study binding events is indicated for each of the five identified binding sites by vertical lines.

at these concentrations. The simulations with 20 O$_2$ molecules were thus chosen as the best balance between good sampling statistics and a reasonable expectation of independent behavior between the O$_2$ molecules and were used for the further analysis of O$_2$ localization and binding.

**Localization of binding sites**

The O$_2$ localization within the ETF was analyzed using the VolMap plugin in VMD,$^{42}$ which was used to measure the local time-averaged concentration of O$_2$ on a 3D lattice with a resolution of 1 Å over the simulations with 20 O$_2$ molecules. The results are shown in Fig. 4a. Regions with increased average concentration, i.e. increased frequency of finding O$_2$ appear as red surfaces in the figure and indicate a likely tendency of O$_2$ molecules to be captured for extended periods during the simulation. Such binding events can also
Figure 4: a: Regions of high probability of finding \( \text{O}_2 \) are indicated by red blobs: in particular, two isosurfaces of the time-averaged \( \text{O}_2 \) concentration during the simulation with 20 \( \text{O}_2 \) molecules are shown as transparent and solid surfaces, where the solid surface indicates the highest concentration. The five best candidate sites for \( \text{O}_2 \) binding are labeled with numbers 1-5, and the distances between flavin and the binding sites in Å are indicated.

b: An example of an \( \text{O}_2 \) molecule bound at binding site 1.

be observed as plateaus in Figs. S1 and S2 in the SI. Five candidate binding sites have been identified and are labeled with numbers 1-5 in Fig. 4, and simulation snapshots with \( \text{O}_2 \) bound in each of the binding sites were extracted. Fig. 4a also shows the distance in Å between flavin and \( \text{O}_2 \) in the extracted snapshots, and Fig. 4b shows an exemplary configuration of \( \text{O}_2 \) bound at binding site 1. Only binding sites within 20 Å of the flavin group of FADH\(_2\) were considered for further investigation, as electron transfer between FADH\(_2\) and \( \text{O}_2 \) is the main focus of the present study, and such an event was considered to be unlikely at greater distances.\(^{51,52}\)

**Binding site characterization**

The binding of \( \text{O}_2 \) at each of the five identified binding sites was studied using 100 independent simulations per binding site, each starting in a bound \( \text{O}_2 \) configuration. Videos of example binding events at binding sites 2 and 5 are available as supplementary video 1 and 2. Figure 5 shows the distribution of the \( \text{O}_2 \) distance from its initial position; here, a
Figure 5: The distribution of distances from the O$_2$ molecule to its initial bound position in the 100 independent simulations for each of the 5 binding sites indicated in Fig. 4. The O$_2$ molecule is initially bound in all simulations. Positions of atoms were collected for each simulation until unbinding occurred, which was defined as the moment when the distance between O$_2$ and its initial position exceeded 10 Å.

A binding site radius of 10 Å was used as an unbinding criterion to terminate the simulations. Based on the obtained distributions in Fig. 5, a distance of 3 Å was then chosen as a refined binding site radius for sites 1-4 for the second step of the analysis.

Binding site 5 did not seem to have a clearly defined binding radius. Fig. 6 shows a volume map of the O$_2$ localization near binding site 5, similar to Fig. 4a, but this time, the map was computed based on the simulations of binding events at binding site 5. There is an approximately spherical region with very high O$_2$ occupancy near the initial position of the O$_2$ molecule, which is not surprising since the 100 independent simulations all start with O$_2$ localized exactly within this small volume. It is, however, more interesting to note the other regions of increased O$_2$ occupancy uncovered in Fig. 6, which indicates that
Figure 6: Localization of the bound O\textsubscript{2} molecule in the simulations of binding events at binding site 5 is indicated by red isosurfaces of the probability distribution of the O\textsubscript{2} position: the red surfaces enclose regions of increased O\textsubscript{2} occupancy. The solid red surface, which indicates a higher probability density value than the transparent surfaces, contains the initial position of O\textsubscript{2}. Binding site 5 is better described as a small cluster of connected binding pockets than a single spherical site.

binding site 5 is rather a small cluster of closely spaced binding pockets. In the interest of keeping the analysis simple, the binding site radius was kept at a value of 10 Å, which is sufficiently large to include all of the high-probability volume shown in Fig. 6.

After refinement of the binding site radius, the O\textsubscript{2} displacement distances were resampled but only until the first unbinding event according to the new radius. Finally, the binding site centers were redefined as the average position of O\textsubscript{2} during the binding event, and the analysis was repeated a third time, leading to the distributions shown as yellow histograms in Fig. 7. The figure shows that the additional peaks for sites 1-4 have been filtered away due to the refined binding site definition criterion. The figure also presents the distributions of the distance of closest approach between O\textsubscript{2} and the flavin group of FADH\textsubscript{2} cofactor in ETF as cyan histograms.
Figure 7: Distributions of the $O_2$ position in the simulations of $O_2$ binding events at each of the five identified binding sites after refinement of binding site centers and radii. The yellow histograms indicate the distribution of the distance between the $O_2$ molecule and the measured binding site center, which is similar to the distributions in Fig. 5, but with refined unbinding criteria (binding site radii) and binding site centers. The cyan histograms represent the sampled distances of closest approach between the flavin group of the FADH$_2$ cofactor and the $O_2$ molecule. The Gaussian fits, Eq. (5), are shown with red lines.
Figure 8: The distance of closest approach between the flavin group and the O₂ molecule is denoted \( r \), while the displacement distance between O₂ and the binding site center is denoted \( R \). Lastly, \( D \) is the distance between flavin and the binding site center. It is assumed in the analysis that a single atom of flavin will be the closest one throughout the binding site region.

To further characterize the binding sites, the amino acids from the ETF protein within 3 Å of the bound O₂ molecule were detected throughout the binding event simulations. The amino acids that were found within this threshold distance during at least 10% of the simulation time before unbinding of O₂, according to the above criteria, were identified as likely relevant for O₂ binding. On average, 5-6 amino acids were thus identified for each binding site, and histograms of the edge-to-edge distance between each of the identified amino acids and the bound O₂ molecule are shown in Fig. S3 in the SI.

In order to make sense of the two types of distance distributions in Fig. 7, a few geometrical considerations are needed; the complete derivation of the following formulas is provided in the SI. Let \( f_R(R) \) be the probability distribution of the displacement distance \( R \) of an O₂ molecule relative to the binding site center (see Fig. 8) and \( f_r(r) \) be the probability distribution of the edge-to-edge distance \( r \) between O₂ and the flavin group of the FADH₂ cofactor. The probability density, \( f_V \), of the O₂ displacement at the binding site per unit volume may be assumed to be a Gaussian function centered at the binding
site center, therefore,
\[ f_V(R) = \frac{1}{\sigma^3 \sqrt{2\pi}} \exp \left( -\frac{R^2}{2\sigma^2} \right), \quad (1) \]
where \( \sigma \) is the spread of the Gaussian distribution. The corresponding radial distribution can be readily obtained by making use of the spherical symmetry of Eq. (1) and taking the spherical shells \( dV = 4\pi R^2 dR \) as volume elements:
\[ f_R(R) = \sqrt{\frac{2}{\pi}} \frac{R^2}{\sigma^3} \exp \left( -\frac{R^2}{2\sigma^2} \right). \quad (2) \]

Even though the probability distribution in Eq. (1) is centered at the binding site center \((R = 0)\), the peak of the radial distribution function in Eq. (2) is expected at a non-zero displacement distance \((R_{\text{max}} = \sqrt{2}\sigma)\) due to the geometry, as is also observed in the yellow histograms in Fig. 7.

The flavin–O\(_2\) distance distribution, \( f_r(r) \), can be estimated, if one assumes that a single atom of the flavin group will be the closest atom to any point with non-vanishing probability of the O\(_2\) position around the binding site, such that \( D \) and \( r \) measure distances from the same atom on the flavin group to the binding site center and to the O\(_2\) molecule, respectively. Assuming a Gaussian profile, the probability density per unit volume can be expressed in spherical coordinates with the flavin atom at the origin:
\[ f_V(r, \theta) = \frac{1}{\sigma^3 \sqrt{2\pi}} \exp \left( -\frac{r^2 + D^2 - 2rD \cos \theta}{2\sigma^2} \right), \quad (3) \]
where \( \theta \) is the polar angle in spherical coordinates (see the derivation in the SI). The radial distribution function can be obtained by employing the cylindrical symmetry and integrating over the spherical surface at constant \( r \):
\[ f_r(r) = \frac{r}{\sigma D \sqrt{2\pi}} \left[ \exp \left( -\frac{(r - D)^2}{2\sigma^2} \right) - \exp \left( -\frac{(r + D)^2}{2\sigma^2} \right) \right]. \quad (4) \]
Table 3: Distances, $D$, between flavin and each of the five studied O$_2$ binding sites according to the fits of the Gaussian model, Eq. (5), to the distributions of distance of closest approach shown in Fig. 7. Also shown are the characteristic binding times, $\tau_B$, obtained by fitting an exponential model, Eq. (6), to the binding time distributions in Fig. 9. The table includes the coefficients of determination for the fits, $R^2$, and the spread, $\sigma$, of the fitted Gaussian functions.

<table>
<thead>
<tr>
<th>Site</th>
<th>$D$ (Å)</th>
<th>$\sigma$ (Å)</th>
<th>$R^2$</th>
<th>$\tau_B$ (ns)</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.63</td>
<td>0.273</td>
<td>0.9870</td>
<td>0.46 ± 0.003</td>
<td>0.9847</td>
</tr>
<tr>
<td>2</td>
<td>2.99</td>
<td>0.583</td>
<td>0.9093</td>
<td>1.21 ± 0.023</td>
<td>0.9512</td>
</tr>
<tr>
<td>3</td>
<td>7.82</td>
<td>0.544</td>
<td>0.9992</td>
<td>1.72 ± 0.016</td>
<td>0.9907</td>
</tr>
<tr>
<td>4</td>
<td>12.04</td>
<td>0.642</td>
<td>0.9981</td>
<td>2.47 ± 0.014</td>
<td>0.9963</td>
</tr>
<tr>
<td>5</td>
<td>13.88</td>
<td>1.085</td>
<td>0.8259</td>
<td>2.41 ± 0.158</td>
<td>0.8147</td>
</tr>
</tbody>
</table>

The first term in the square brackets is a Gaussian function centered at the binding site center, while the second term is vanishingly small, except when $D \ll \sigma$ — essentially when the binding site overlaps with the atom on the flavin, which is not the case for the cyan distributions shown in Fig. 7. To the first order, the prefactor $r$ leads to a small shift in the position of the Gaussian peak by $\frac{\sigma^2}{2D}$; approximating $f_r(r)$ as:

$$f_r(r) \approx \frac{1}{\sigma \sqrt{2\pi}} \exp\left(\frac{\sigma^2}{2D^2}\right) \exp\left(-\left(\frac{r - (D + \frac{\sigma^2}{2D})}{2\sigma^2}\right)^2\right).$$

(5)

Fits of Eq. (5) are shown in Fig. 7 as red curves, and the fitted distances and spreads, $D$ and $\sigma$, are listed in Table 3. Binding sites 1, 3 and 4 appear to be well described by the Gaussian model, while sites 2 and 5 have a higher divergence. For site 5, this is not surprising, as this binding site consists of a small cluster of binding pockets, see Fig. 6. For site 2, the skew distribution is likely arising due to a non-spherical binding site, but could also be partially attributed to a violation of the assumption that only a single atom of the flavin is the closest one for the entire binding site region, as site 2 is rather close to the flavin group.
**Binding time analysis**

The binding sites are small volumes surrounded either fully or partially by potential barriers, such that an oxygen molecule would be trapped there for some time, but since an oxygen molecule was able to enter the binding site in the first place, it will eventually leave the binding site again. Assuming that unbinding is a Poisson process, i.e. that at any given time, a bound O$_2$ molecule has an equal probability of unbinding, no matter how long it was already bound, the probability of an O$_2$ molecule staying bound for a given duration will decay exponentially with some characteristic time, $\tau_B$. For the $N_0 = 100$ independent simulations per binding site starting with a bound configuration, the number of simulations expected to still have oxygen bound at time instance $t$ is, therefore, distributed according to an exponential model:

$$N(t) = N_0 \exp \left( -\frac{t}{\tau_B} \right).$$

(6)

The distribution of bound oxygen molecules is illustrated in Fig. 9 together with the fits to the exponential model in Eq. (6). The characteristic binding times for the binding sites are listed in Table 3. Binding sites 1–4 fit the exponential model reasonably well, which is consistent with them being simple, approximately spherical sites. The more complex shaped site 5, on the other hand, is poorly described by the model, and in particular has a long tail of O$_2$ molecules, which remain bound longer than the initial exponential decay would predict. This indicates that when the entire region indicated in Fig. 6 is included, binding site 5 has the potential to trap O$_2$ molecules for significantly longer times than the 2.4 ns indicated in Table 3, but the bound molecules will be able to move within the somewhat larger region of confinement compared to binding sites 1–4.
Figure 9: Number of bound O\textsubscript{2} molecules as a function of simulation time in the 100 independent simulations with bound O\textsubscript{2}, which were studied for each of the five binding sites until unbinding occurred. The red curves indicate the fits of the exponential model in Eq. (6). The fitted characteristic binding times are listed in Table 3.

**Potential for radical pair formation**

Binding of molecular oxygen in the ETF enzyme with a fully reduced FADH\textsubscript{2} cofactor leads to the potential of radical pair formation through an electron transfer event between FADH\textsubscript{2} to O\textsubscript{2} to produce an FADH\textsuperscript{•−}–O\textsubscript{2}\textsuperscript{−} pair,\textsuperscript{53} see Fig. 10, which in turn leads to the potential release of harmful ROS in mitochondria. To form a radical pair, FADH\textsubscript{2} would need to release both an electron and a proton and form a neutral FADH\textsuperscript{•} radical. While the electron would be transferred to a bound O\textsubscript{2} molecule, the proton would likely
Figure 10: Reduced flavin activates O₂ to produce a spin-correlated radical pair between flavin semiquinone and O₂⁻. A branching point exists that can produce O₂⁻ or H₂O₂, depending on the spin state of the radical pair.

be accepted by a nearby amino acid from the protein, similarly to other proton-coupled electron transfer processes common in biological charge transport mechanisms.⁵⁴–⁵⁶ Fig. 11 shows a δ-protonated histidine, α-HIS286, with a small distance to flavin, stabilized by a hydrogen bond. The histidine could potentially accept the released proton in a combined electron and proton transfer process to become biprotonated, e.g. through rotation of the imidazole ring in concert with the proton transfer.

Taking a closer look at the identified O₂ binding sites and their potential for radical pair formation, the local environment at each binding site is illustrated in Fig. 12, including the
Figure 11: A histidine very close to the flavin group of FADH$_2$ could potentially accept a proton in a reaction involving electron transfer to O$_2$, i.e. the first reaction in the scheme depicted in Fig. 10.

Among the studied binding sites, site 5 comes out as the most interesting with respect to superoxide formation. Firstly, the positively charged $\beta$-ARG12 in the local binding environment of site 5 will lower the free energy of the product state of the electron transfer reaction that would produce superoxide. The positively charged $\alpha$-ARG249 at binding site 2 has significantly less contact with the bound O$_2$ molecule than $\beta$-ARG12 at binding site 5, as can be seen from Fig. S3. Hence, there might be a large driving force for the electron transfer from the FADH$_2$ cofactor at site 5, and the binding time of a produced superoxide anion could be significantly increased compared to binding sites 1–4. This is in particular in contrast to binding sites 3 and 4, which have negatively charged amino acids in the local binding environment. The local superoxide binding affinities will be studied in more detail in a follow-up study, where the same methodology used here to study the dynamics
Figure 12: The local environment at the five identified binding sites. Only amino acids with an atom within 3 Å of $O_2$ are considered. Negatively charged amino acids are highlighted with red color, positively charged with blue, while polar and non-polar neutral amino acids are indicated with green and gray, respectively. Note that GLU165 in binding site 4 is protonated here and therefore shown as neutral. Amino acids labels are green for the alpha subunit and purple for the beta subunit of ETF.

of oxygen in the ETF protein could be applied instead to superoxide. Secondly, site 5 is also of particular interest since the radicals of a putative radical pair between flavin and $O_2^{-}$ formed at site 5 would be far apart, such that exchange and magnetic dipole-dipole interactions would be much smaller than at each of the other binding sites, which is a necessary condition for weak magnetic fields to have any impact on the radical pair.\textsuperscript{57}

While the characteristic binding times are fairly short (less than 3 ns), binding site 5 shows evidence of a sub-population of $O_2$ molecules, which remain bound for significantly longer times, increasing the probability of superoxide production. The large separation will, on the other hand, also make electron transfer events less frequent, but electron transfers between cofactors with similar separation distances within a protein are know to occur.\textsuperscript{52,58}
The proposed redox mechanism of the $O_2$–FAD reaction comes from studies with purified FAD in solution by others. However, ETFs are soluble heterodimeric FAD-containing proteins that are found in all kingdoms of life. The formation of a caged radical pair in these enzymes is a mandatory step but is far from being established experimentally. In accordance with that, our results indicate that there are five regions of high probability of binding $O_2$, and none of them are close to C(4a). These results open new doors to establish the mechanism of formation of ROS from $O_2$ and ETFs.

**Conclusion**

The present investigation focuses on modeling the first step of a possible radical pair formation within the mitochondrial ETF enzyme: namely, the binding of molecular oxygen within the protein, leading to the opportunity of electron transfer events between the FADH$_2$ cofactor and $O_2$. The computational studies revealed a number of possible $O_2$ binding sites in the ETF protein with a fully reduced FADH$_2$ cofactor, and through statistics from extensive simulations and mathematical modeling, the shapes and binding times of the sites were characterized. The complex-shaped binding site 5 has several features that makes it interesting as a candidate site for further studies on the possibility of radical pair formation: the distance between the FADH$_2$ cofactor and the binding site would likely lead to sensitivity of the formed radical pair to rather weak external magnetic fields, and two positively charged amino acids in the binding environment could stabilize superoxide at the protein. The measured characteristic binding times were generally rather short ($< 3$ ns), but may nevertheless be sufficient for an efficient electron transfer. For site 5, in particular, there appears to be a secondary population of $O_2$ molecules remaining bound for a longer time, which further adds to its relevance for a potential radical pair formation mechanism. The possibility for formation of radical pairs within the ETF protein is significant, as it
means that the release of ROS could be controlled by the coherent spin dynamics of a radical pair reaction.

The methodology presented here is not developed uniquely for molecular oxygen, but could readily be applied to study the behavior of other small molecules of physiological interest, such as CO or NO, within a protein environment, and the molecular dynamics approach should be able to capture differences due to charge and polarity of the studied molecules. In particular, the approach could be used in a subsequent study to chart the dynamics and binding of superoxide, once produced through oxidation by FADH$_2$ in order to complete the picture and measure superoxide binding times. A follow-up study could also attempt to model the spin dynamics of a putative radical pair and show evidence of the requisite lifetimes to support coherent transition pathways in the ETF. Briefly, one of the first steps in ROS production is the one-electron reduction of O$_2$ that results in the formation of O$_2^*$ as shown in Fig. 10. In this general reaction scheme, an electron is transferred from reduced flavin to triplet O$_2$, which results in O$_2^*$ and a flavosemiquinone (FADH*) spin-correlated caged radical pair.$^4$ A ROS branching point exists where the reaction has two separate pathways, which – under specific magnetic field conditions – can be controlled by singlet-triplet interconversion.$^{59-61}$ At the most fundamental level, one should investigate how such a radical pair reaction can be controlled by perturbing spin interconversion through applied optimal magnetic fields to control the outcome of ROS production in the ETF enzyme. Crucially, this will lead to predictive power of the computational model, which could, possibly along with in silico mutational analysis (e.g. β-D128N and β-R191C known to influence ROS production$^{10}$), allow for experimental validation of the approach.
Supporting Information

Plots of the edge-to-edge distance between the flavin group and the nearest $O_2$ molecule as a function of simulation time in the simulations with 10-40 $O_2$ molecules; mathematical derivation of radial distance distributions; histograms of the distance between a bound $O_2$ molecule and nearby amino acids at each of the five binding sites during the binding event simulations; partial charges and force-field parameters for FADH$_2$ obtained through quantum chemical calculations and fitting using ffTK (PDF). Supplementary videos 1 and 2 showing example binding events at binding site 2 and 5, respectively.

Acknowledgements

The authors thank the Lundbeck Foundation, the Independent Research Fund Denmark, Deutsche Forschungsgemeinschaft (GRK1885), the Volkswagen Foundation (Lichtenberg professorship to IAS) and the Air Force Office of Scientific Research (grant no. FA9550-17-1-0458) for financial support. Computational resources for the simulations were provided by the DeiC National HPC Center, SDU. The authors thank Robert Usselman for fruitful discussions.

References


