Simulating substrate recognition and oxidation in laccases: from description to design

Maria Fátima Lucas¹, ²‡*, Emanuele Monza¹‡, Lise J. Jørgensen, ³ Heidi A. Ernst, ³ Klaus Piontek, ⁴ Morten J. Bjerrum, ³ Ángel T. Martinez⁵, Susana Camarero⁵, Victor Guallar¹, ⁶*  

¹Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, Jordi Girona 29, E-08034 Barcelona, Spain ²- Anaxomics Biotech, Balmes 89, E-08008 Barcelona, Spain ³- Department of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, C. ⁴- University of Freiburg, Albertstraße 21, D-79104 Freiburg im Breisgau ⁵- Centro de Investigaciones Biológicas, CSIC, Ramiro de Maeztu 9, E-28040 Madrid, Spain. ⁶-ICREA, Passeig Lluís Companys 23, E-08010 Barcelona, Spain.

ABSTRACT

To meet the very specific requirements demanded by industry, proteins must be appropriately tailored. Engineering laccases, to improve the oxidation of small molecules, with applications in multiple fields, is however a difficult task. Most efforts have concentrated on increasing the redox potential of the enzyme but in recent work we have pursued an alternate strategy to engineering these biocatalysts. In particular, we have found that redesigning substrate binding at the T1 pocket, guided by in silico methodologies, to be
a more consistent option. In this work, we evaluate the robustness of our computational approach to estimate activity, emphasizing the importance of the binding event in laccase reactivity. Strengths and weaknesses of the protocol are discussed along with its potential for scoring large number of protein sequences and thus its significance in protein engineering.

INTRODUCTION

Laccases (EC 1.10.3.2) are multicopper oxidases whose catalytic core is organized in two copper centers: the T1 site, where substrates are oxidized, and the trinuclear cluster, where molecular oxygen is reduced to water. The broad specificity of these proteins, the use of oxygen as final electron acceptor and its conversion into water make them ideal candidates for sustainable industrial processes.1 On the other hand, the T1 copper's redox potential (E_{T1} < 0.8 V) constitutes an upper limit to their application. Although mediators are successfully used to extend their activity toward high redox potential substrates,2 trying to broaden the chemical space of these proteins by increasing their redox potential is a challenging strategy.3 The observation of a linear dependence between the logarithm of the specificity constant (k_{cat}/K_M) and the one-electron redox potential difference between laccase's T1 copper site and phenolic substrates has for a long time fueled this approach.4 However, exceptions can be found in the literature, such as the oxidation of syringaldazine by Myceliophthora thermophila (MtL, E_{T1}^o=0.46 V), Rhizoctonia solani (E_{T1}^o=0.73 V) and Pycnoporus cinnabarinus (PcL, E_{T1}^o=0.79 V), which was shown to be more favorable in the enzyme with the lowest redox potential.5,6 In fact, in addition to the redox potential difference, the oxidation rate is also controlled by the binding event,7 which directly
influences the electron transfer (ET) driving force (shifting the energy levels of the substrate), the electron coupling (determining the donor-acceptor distance and their relative orientation) and the reorganization energy (modulating substrate's solvent accessible area and active site preorganization). To advance in this matter we have used a recently developed computational protocol\(^7\) that combines the protein energy landscape exploration (PELE) software,\(^8,9\) to map the laccase-substrate recognition process at the T1 copper (Cu) site, with quantum mechanics/molecular mechanics (QM/MM) techniques, to score enzyme reactivity.

In this work, we systematically apply Monza et al.’s methodology\(^7\) to study the oxidation of five substrates: syringaldazine (SGZ – see Scheme 1) and four phenols with general formula 4X-phenol (X = OH, OMe, Me and Cl; named from this point on as 4OH, 4MeO, 4Me and 4Cl, respectively) by two fungal laccases: a high redox potential laccase from the basidiomycete PcL (\(E^\circ_{T1} = 0.79\) V) and a medium redox potential laccase from the ascomycete MtL (\(E^\circ_{T1} = 0.46\) V)

Scheme 1. Chemical structure of all studied substrates.
The manuscript is divided in two main sections: First we investigate the unusual reactivity profile of syringaldazine by MtL and PcL. Results confirm earlier findings on the importance of the binding event in enzyme activity irrespective of the enzyme’s redox potential. Then, we perform a systematic study of the oxidation of a series of related substrates by MtL and PcL. These substrates were chosen because their reactivity correlates with their driving force, approximated as redox potential difference between electron acceptor and donor, and to minimize potential changes in reactivity due to substrate size and/or bulky substituents. The importance of a correct description of the binding site (for example the explicit inclusion of crystal waters) and how clustering methods are required to identify the most reactive positions are also discussed.

**COMPUTATIONAL DETAILS**

**Systems setup.** The crystal structures of PcL (2XYB.pdb) and MtL (provided by Prof. Bjerrum, to be published) were prepared with Protein Preparation Wizard, \(^{11}\) PROPKA utility \(^{12}\) and the H++ web server \(^{13}\) to determine the protonation state of all titratable groups at pH 5. Five substrates were prepared: syringaldazine (SGZ) and the four phenols derivatives 4OH, 4MeO 4Me and 4Cl. The redox potential of these phenols is 0.48; 0.66; 0.79; 0.90 V, respectively. \(^{4}\) All substrates were fully optimized with the density functional M06\(^{14}\) with the 6-31G** basis set in an implicit solvent (modeled through the Poisson-Boltzmann equation) and the atomic charges, computed at the same level of theory by fitting the molecular electrostatic potential, were used in the following force field based PELE simulations.
Laccase-substrate recognition process. To determine the different binding modes of all substrates in the T1 Cu site of MtL and PcL, PELE simulation were performed. PELE, which has been successfully applied to study enzymes,\textsuperscript{7,10,15-25} is a Monte Carlo algorithm that combines protein backbone (elastic network model\textsuperscript{26}) and ligand displacements (roto-translations), followed by side chain repacking\textsuperscript{27} and all atom minimization. The new positions are then accepted or rejected through a Metropolis test based on energy differences computed using an OPLS force field\textsuperscript{28} and a surface-generalized Born implicit continuum solvent.\textsuperscript{29} In this study, the substrates are manually placed close to the entrance of the T1 site and are then free to move in a region 20 Å within the Cu atom. An example of such a trajectory can be seen in supplementary material. The results from local conformational searches for SGZ in PcL and MtL (96 independent trajectories produced in 48h for each system), taken as an example, are visualized in interaction plots (Figure 1). These contain all accepted minima’s interaction energies against the distance of the center of mass of the ligand to the Cu atom at the T1 site. Interaction energies are computed as 
\[ E_{\text{INT}} = E_{\text{PS}} - E_P - E_S \] where PS refers to the protein (P) substrate (S) complex.
Figure 1. Interaction energies vs. distance between the center of mass of SGZ and the T1 Cu atom in: A) MtL and B) PcL. The different colors indicate independent single processor simulations.

Structures of interest are then randomly selected up to 10 Å of the Cu atom and within 5 kcal/mol of the lowest energy value (red rectangle in Figure 1).

**QM/MM scoring.** The selected structures obtained in the previous PELE simulations were used to estimate the amount of spin density transferred from the substrate to the Cu site. For this, we have used a QM/MM scheme to model the whole laccase-substrate complex which employs different levels of theory to describe the system. QSite\textsuperscript{30} was used including in the quantum region the entire Cu site (with equatorial and axial ligands) as well as the substrate. The M06-L\textsuperscript{14} density functional with the lacvp* basis set was used (LANL2DZ effective core\textsuperscript{31} for the Cu atom and 6-31G* for the rest of the atoms). The
remaining part of the protein was modeled with molecular mechanics through an all-atom OPLS force field. A five step geometry QM/MM optimization was carried out and Mulliken populations were computed to characterize the fraction of spin density transferred from the substrate to the Cu site. Previous experimental studies have shown that spin densities correlate well with the ET driving force by establishing if an unpaired electron is energetically more stable on the donor or acceptor’s molecular orbitals.\textsuperscript{32} This quick optimization has been proven to be sufficient to obtain converged spin densities. In laccases, moreover, the calculated spin population has been shown to be largely invariant to changes in density functional, basis set and initial guess.\textsuperscript{7}

**Molecular dynamics.** To ensure that two substrate molecules can coexist in the active site, molecular dynamics simulation were run for the PcL-4OH system, starting from the configuration depicted in Figure S9. The ternary complex was solvated with a 10 Å buffer of waters in an orthorhombic box, neutralized and 0.15 M NaCl was added. After equilibration (default settings), 8x5 ns production runs at constant temperature (300 K) and pressure (1 atm) were performed with Desmond.\textsuperscript{33} The OPLS-2005 force-field\textsuperscript{28} and the SPC explicit water model\textsuperscript{34} were used. The copper centers were generated with the hetgrp_ffgen utility of Schrödinger, using classical charges and crystal structure geometries. The temperature was regulated with the Nosé–Hoover chain thermostat\textsuperscript{35} with a relaxation time of 1.0 ps, and the pressure was controlled with the Martyna–Tobias–Klein barostat\textsuperscript{36} with isotropic coupling and a relaxation time of 2.0 ps. The RESPA integrator\textsuperscript{37} was employed with bonded, near, and far time steps of 2.0, 2.0, and 6.0 fs,
respectively. A 9 Å cutoff was used for non bonded interactions together with the smooth particle mesh Ewald method.\textsuperscript{38}

RESULTS AND DISCUSSION

Syringaldazine oxidation by MtL and PcL.

Although unusual, there are some examples in the literature where lower redox potential laccases display better $k_{\text{cat}}$ than their higher redox potential counterparts.\textsuperscript{39} In particular, the oxidation of syringaldazine, by MtL ($k_{\text{cat}}=1100 \text{ min}^{-1}$) has been shown to display improved kinetics over PcL’s ($k_{\text{cat}}=180 \text{ min}^{-1}$).\textsuperscript{5} Here, aiming at understanding these differences, we have investigated the oxidation of SGZ by MtL and PcL using computational tools. For this, 96 independent trajectories were produced, with PELE, to establish the most favorable laccase-SGZ binding modes. Then, 20 structures were randomly selected with an energy-distance filter (see Methods) and QM/MM calculations performed to assess the fraction of spin density transferred from the substrate to the enzyme.

The interaction energy profiles for SGZ diffusion in MtL and PcL are considerably different (Figure 1) and it is clear that binding is more favorable (with a funnel-like profile and lower interaction energies) in MtL than PcL. Additionally, minima in MtL are closer to the electron acceptor and better protected from the solvent (in theory, lowering the ET coupling and reorganization energy, respectively) as seen in Figure S1. More importantly, SGZ shows a better orientation for both electron and proton transfer in MtL (Figure 2).
The differing binding of SGZ in MtL and PcL derives from differences in the T1 pocket. MtL contains a large loop involving residues 445 and 468 hosting Asn454 that anchors the substrate in the position seen in Figure 2A. PcL, on the contrary, has a much smaller loop of only 8 residues (opposed to the 23 in MtL) involving residues 408 to 416. Another significant difference is the shorter loop hosting residues 332 to 336 in PcL (364 to 371 in MtL) that creates a favorable environment to dock the polar hydroxyl group of SGZ (Figure 2B). Finally, Ala192, which helps positioning the substrate in the correct orientation in MtL’s active site, is located at the end of a loop that is preceded by an α-helix beginning in residue 179 which is not present in PcL (left side of Figure 2A). In addition to the considerable differences seen both in the interaction energy profiles, donor-acceptor distances, solvent exposure and binding orientation, QM/MM calculations also indicate higher spin transfer from SGZ to MtL than in PcL (Figure 3).
Figure 3. Distribution of SGZ spin densities in MtL and PcL in boxplot representations. In each distribution of data we have: minimum, first quartile, median, third quartile, and maximum. Each sphere is one computed spin density.

The computational results put in evidence the fact that counterintuitive differences in SGZ’s oxidation by these two enzymes are highly related to the binding event. In MtL, SGZ favors configurations hydrogen bonded (H-bonded) to the first coordination His508 (facilitating ET) and simultaneously interaction with Glu235 where proton transfer is expected to occur.\textsuperscript{40} This will ensure optimal electronic coupling (shorter donor-acceptor distance) and ET driving force (higher substrate spin density, likely due to the proximity to the catalytic Glu235). In PcL, however, SGZ interacts with the backbone of residues Gly334 and Asn336 and with the side chain of Phe265 anchoring the substrate about 3 Å away from any of the T1 ligands. Indeed, Phe265 has been described to mark the boundary to the entrance channel to T1 site in the high-redox potential laccase from \textit{Trametes versicolor}, 1KYA, with 66% sequence identity with PcL.\textsuperscript{41} In this higher redox enzyme,
electron transfer must thus occur through a free space jump (as none of the coordinated residues is close enough to the substrate) or alternatively using other residues (which necessarily imply a longer electron transfer path). These results advocate for improved kinetics in MtL, which is in good agreement with experimental observations.

4X-Phenol oxidation by MtL and PcL.

Preliminary PELE simulations for the 4X-phenols with PcL showed different binding modes for each laccase-substrate system. However, QM/MM calculations displayed poor correlation between the amount of spin density computed on each substrate (20 structures selected) and the substrate’s redox potential (Figure S2 in supporting information). Further inspection of the most favorable minima evidenced an important cluster of structures, for all complexes, that included interaction with His456 (first coordination sphere of the Cu T1 site) but not with the catalytic aspartic acid, which is expected to be the proton acceptor (Figure S3A).

These structures exposed a region in PcL’s binding pocket with a volume comparable to that occupied by a water molecule. Further inspection of the available laccases’ crystal structures that contained a phenolic substrate showed, for example, that in *Melanocarpus albomyces* laccase in addition to 2,6-dimethoxyphenol it has a crystal water molecule precisely in the position identified in the simulations (Figure S3B). For this reason, we have repeated all calculations (PELE + QM/MM), involving the phenols, including an extra water molecule in this position. For each laccase-substrate pair, 240 independent 48h trajectories were produced (interaction energy profiles in Figure S5) and 50 complex structures were randomly selected for QM/MM scoring. Computed spin densities for all
systems (depicted in Figures S6 and S7) offer a qualitative picture where we can see the overall improved oxidation of compound 4OH over 4Cl and that complexes with PcL display, in average, higher spin transfer than with MtL. Next we identified the site with maximum substrate spin density, for each compound, by clustering techniques. For this, computed spin densities (50 for each system) were grouped using k-means and k-medoids and the obtained clusters (details can be found in SI Table S8 and Figures S9 and S10) were then visually inspected. For both laccases, the highest spin density cluster is also the one with optimal catalytic contacts. The substrate is H-bonded to the catalytic histidine (first electron acceptor, coordinated to the T1 copper) and to the water molecule present in the active site. This is in turn hydrogen bonded to the catalytic Asp, the final proton acceptor (Figure S4).

It follows that such binding mode, besides from an excellent driving force (most likely due to the proximity of the negatively charged catalytic base), provides both optimal electron tunneling and proton abstraction. If only this cluster of best oxidation position is taken into account, for each system, a good correlation (above 95%) between the computed spin density and the logarithm of the experimental specificity constants is obtained (Figure 4).
Figure 4 - Computed average spin densities in the sites identified as the best oxidation positions in the T1 copper site for MtL and PcL vs. experimental specificity constants\textsuperscript{4} for MtL (blue) and TvL (cyan).

Since specificity constants are not available for PcL we compared with another high redox potential enzyme from \textit{Trametes villosa} (E\textsubscript{o,T1}=0.78 V) which shares 70% identity (homology modelling shows an identical ligand binding site to PcL with only Phe164 replaced by an Ile in the binding pocket entrance). These small differences may be responsible for the lower correlation found for this system. To support the hypothesis that, at high substrate concentration, the average spin density at the best oxidizing site is correlated with experimental specificity constants, we must ensure that occupation of the best site is possible when other substrate molecules are simultaneously bond at neighbor
positions. For this, we have performed eight independent 5 ns molecular dynamics simulations and show that two ligands can, in fact, concomitantly exist for at least 2 ns around their initial positions (Figures S11 and S12). Furthermore, QM/MM calculations confirm that the best site is the one oxidized by the enzyme independently of the second substrate molecule. This hypothesis is expected to be valid in systems where multiple positions can be simultaneously occupied and where the presence of a second ligand does not affect the oxidation of the first (for example charged substrates that change the electrostatic environment of the substrate can affect the ET).

In any case, results confirm the potential of the in silico protocol to discriminate laccase reactivity. Qualitatively scoring enzyme activity is shown to be straightforward and the method is robust enough to capture even unusual trends as with the oxidation of SGZ by PcL and MtL. The present study further validates a general methodology already used to explain directed evolution experiments\(^7,10\) and computer-aided rational design of laccases.\(^15\)

In the latter, the oxidation of a poorly reactive molecule was improved by introducing a negatively charged residue to increase its oxidability. This was achieved by tuning the binding event (PELE) and enhancing the spin density (QM/MM). In the case of determining the oxidation activity toward different substrates it is clear that the problem becomes much more complex. A correct description of the binding site, namely with the inclusion of essential water molecules, is fundamental. Furthermore, we find that when multiple binding modes exist, if they are non-competitive, then the oxidation is mainly driven by the best oxidation position and clustering methods can help finding it.

An identical approach (PELE + QM/MM) was also recently used to design a stable manganese peroxidase\(^20\) which indicates that the methodology benchmarked here can be
easily extended to other oxidoreductases. Nevertheless, and as stated previously, our aim is
to develop cost effective and reliable computational strategies for protein engineering. In
this context we wish to move away from rational design (where a few variants are designed
and validated experimentally) and instead provide an exhaustive protein engineering
protocol with reduced human intervention. The goal is to either compute a large set of point
mutations and test the most promising experimentally or identify amino acids or regions of
the protein where directed evolution experiments should focus their effort. In this direction,
we have recently shown how the protocol benchmarked here can be used for systematic
sequence space search. In the work by Giacobelli et al., we have used PELE to explore
the binding in the parental type. After identifying the most reactive laccase-substrate
conformation, over 400 mutants were evaluated with a quick (less than 1 hour in single
processor per mutant) PELE protocol. These calculations can be easily done in less than
one day with 100 CPUs, providing an inexpensive alternative to laboratory directed
evolution. In fact, we have shown that while QM/MM scoring (which is the most time
consuming step) provides a reliable assessment it is possible to qualitatively measure
catalytic improvement, using only structural and force field based parameters such as
donor-acceptor distance, catalytic contacts and interaction energy. Final QM/MM scoring,
which takes 2-4 hours on 4 CPUs, can easily be limited to re-rank the best 10s-100s
variants thus making this protocol appropriate for academic and industrial labs, particularly
in a first approach to engineering a protein.

In conclusion, we demonstrate the robustness of the computational method here
benchmarked to accurately predict the oxidation ability of laccases towards small
molecules. We discussed its potential to quickly score the activity of a large number of
protein sequences toward the oxidation of specific substrates and thus the importance of *in silico* methodologies in protein engineering.

ASSOCIATED CONTENT

Supporting Information. File S1 including an example of PELE simulation (movie) and File S2 including PELE simulations for SGZ with MtL and PcL (SASA plots vs. distance) (**Figure S1**), spin distribution for 4X-phenols preliminary tests without a water molecule (**Figure S2**), PELE simulations for 4X-phenols with MtL and PcL (interaction energy plots vs. distance), (**Figure S3**), Binding pocket comparison between PcL and MaL, (**Figure S4**), Binding positions for 4Cl-phenol in PcL, (**Figure S5**), substrate spin density boxplots for 4X-phenols with PcL and MtL (**Figures S6 and S7**), cluster analyses (**Table S8**, and Figures S9 and S10), and example of two 4OH minima in PcL (**Figure S11**) are available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: victor.guallar@bsc.es

*E-mail: fati.lucas@gmail.com

Author Contributions

‡These authors contributed equally.

Funding Sources
This study was supported by the OxiDesign (CTQ2013-48287-R) and NOESIS (BIO2014-56388-R) Spanish projects and the INDOX (KBBE-2013-7-613549) EU-project. The authors have no conflict of interest to declare.

**ABBREVIATIONS**

$E^o$, redox potential; PcL, *Pycnoporus cinnabarinus* laccase; MtL, *Myceliophthora thermophila* laccase; SGZ, syringaldazine; 4Cl, 4Cl-phenol; 4Me, 4Me-phenol; 4MeO, 4MeO-phenol; 4OH, 4OH-phenol; PELE, protein energy landscape exploration; QM/MM, quantum mechanics/molecular mechanics.

**REFERENCES**


(9) Cossins, B. P.; Hosseini, A.; Guallar, V. Exploration of Protein Conformational...


(35) Nosé, S.; Shuichi, N. A Unified Formulation of the Constant Temperature Molecular


