Accurate Stabilities of Laccase Mutants Predicted with a Modified FoldX Protocol

Christensen, Niels Johan; Kepp, Kasper Planeta

Published in:
Journal of Chemical Information and Modeling

Link to article, DOI:
10.1021/ci300398z

Publication date:
2012

Document Version
Publisher’s PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Accurate Stabilities of Laccase Mutants Predicted with a Modified FoldX Protocol

Niels J. Christensen and Kasper P. Kepp*

Technical University of Denmark, DTU Chemistry, Kemitorvet 206, Kongens Lyngby, DK 2800 Denmark

ABSTRACT: Fungal laccases are multicopper enzymes of industrial importance due to their high stability, multifunctionality, and oxidizing power. This paper reports computational protocols that quantify the relative stability (ΔΔG of folding) of mutants of high-redox-potential laccases (TvLIIIb and PM1L) with up to 11 simultaneously mutated sites with good correlation against experimental stability trends. Molecular dynamics simulations of the two laccases show that FoldX is very structure-sensitive, since all mutants and the wild type must share structural configuration to avoid artifacts of local sampling. However, using the average of 50 MD snapshots of the equilibrated trajectories restores correlation (r ~ 0.7–0.9, r² ~ 0.49–0.81) and provides a root-mean-square accuracy of ~1.2 kcal/mol for ΔΔG or 3.5 °C for T₅₀, suggesting that the time-average of the crystal structure is recovered. MD-averaged input also reduces the spread in ΔΔG, suggesting that local FoldX sampling overestimates free energy changes because of neglected protein relaxation. FoldX can be viewed as a simple “linear interaction energy” method using sampling of the wild type and mutant and a parametrized relative free energy function: Thus, we show in this work that a substantial “hysteresis” of ~1 kcal/mol applies to FoldX, and that an improved protocol that reverses calculations and uses the average obtained ΔΔG enhances correlation with the experimental data. As glycosylation is ignored in FoldX, its effect on ΔΔG must be additive to the amino acid mutations. Quantitative structure–property relationships of the FoldX energy components produced a substantially improved laccase stability predictor with errors of ~1 °C for T₅₀ vs 3–5 °C for a standard FoldX protocol. The developed model provides insight into the physical forces governing the high stability of fungal laccases, most notably the hydrophobic and van der Waals interactions in the folded state, which provide most of the predictive power.

INTRODUCTION

Understanding and predicting protein stability at nonphysiological pH, T, solvent, or ionic strength is of substantial interest in industrial processes where extremophiles are increasingly employed for their efficient and environmentally friendly catalysis.⁴⁻⁶ Industrial conditions often deviate markedly from those of natural habitats, leading to a need for prolonged protein lifetime (i.e., reduced cost)⁴ in nonaqueous solvents, at extreme pH, or at elevated temperatures that increase turnover for industrial processes while often reducing microbial contamination.⁵,⁶ Also, extremophiles are of academic interest in the search for molecular fundamentals of protein folding and stability.⁷

Standard laboratory approaches toward enhancing protein stability include site-directed mutagenesis, directed evolution, and various semirational approaches such as structure-based design and phylogenetic methods that induce sequence traits of known extremophiles or evolutionary ancestors.⁸ In particular, directed evolution using random mutagenesis under specific selection pressures has been used successfully in many studies to amplify desired properties of proteins.⁹,¹⁰ The development of superior industrial enzymes may be advanced by rational methods.¹¹,¹² Although not widely appreciated, rational methods may, if they leap beyond single-point mutations, allow the identification of new desirable proteins relatively far from the wild type (WT) in sequence space. Such proteins are experimentally inaccessible due to the large number of mutant generations required and to the correlation between amino acids, which render beneficial pair substitutions extremely unlikely unless directly introduced (the most simple examples being cystines or salt bridges introduced in a single generation). These desirable, remote islands are separated in sequence space by large oceans of less stable and functional “dead ends”, as e.g. illustrated by the folded but nonfunctional de novo designed protein only 50% similar to its Trametes versicolor laccase scaffold,¹³ a protein studied in this work. Thus, the future success of rational, computational protein evolution depends directly on the ability to predict stability far from the WT in sequence space.

Random errors in state-of-the-art experimental estimates of in vitro protein stabilities from careful protocols that efficiently reduce systematic errors have upper limits in correlation coefficients r of ~0.86 (r² ~ 0.74)¹⁴ between experimental and computed relative free energies of protein folding for mutants vs WT (called herein ΔΔG), defined as

Received: August 22, 2012  
Published: October 27, 2012
\[ \Delta \Delta G = \Delta G_M - \Delta G_{WT} \]
\[ = -RT \ln \left( \frac{K_M}{K_{WT}} \right) \]
\[ = -RT \ln \left( \frac{[F]_M/[U]_M}{[F]_{WT}/[U]_{WT}} \right). \]

with typical experimental standard errors of \( \Delta \Delta G \) of \( \sim 0.5 \) \( \text{kcal/mol} \). Here, \( \Delta \Delta G_M \) and \( \Delta \Delta G_{WT} \) are the free energies of folding for the mutant and WT proteins, \( K_M \) and \( K_{WT} \) are the effective two-state folding equilibrium constants, and \([F]\) or \([U]\) denote concentrations of folded (F) or unfolded (U) protein at equilibrium. A substantial systematic error component is removed by considering only stabilities relative to the WT. A qualitative benchmark criterion for protein stability prediction is the percent of stabilizing and destabilizing mutants predicted correctly, typically at best 80%.

A systematic error component is removed by considering only stabilities relative to the WT. A qualitative benchmark criterion for protein stability prediction is the percent of stabilizing and destabilizing mutants predicted correctly, typically at best 80–90%. Most protocols do not probe the protein’s stability in its natural environment and state, both in terms of pH, \( T \), ionic strength, solvent type, and post-translational modification, including the possibility of several resting, active, and redox states. Still, the unmodified, folded protein stability under well-defined \( \text{in vitro} \) conditions is the standard measure of stability that is also investigated here.

Currently, computational screening of protein mutants can be achieved by software such as, e.g., CUPSAT, FoldX, Dmutant, FoldX, Mutant2.0, and PoPMuSiC. In contrast to computationally intensive methods such as free energy perturbation or thermodynamic integration, these programs attempt to circumvent the structure-sampling problem and thus predict stability on a time scale suitable for high-throughput use. Such programs have been recently evaluated. While \( r \) values up to \( \sim 0.9 \) \( (r^2 \sim 0.81) \) are reported in validations based on single-site mutations in similar proteins from the original papers, a typical \( r \) of \( \sim 0.26-0.59 \) \( (r^2 \sim 0.07-0.35) \) or \( \sim 60\% \) qualitative predictability was found upon external validation on more diverse proteins, showing that, as for other empirical potentials, protein stability predictors are usually of little use outside their parametrization range, even for single-point mutations. Also, computational prediction of single-point mutant stability is hardly useful considering today’s highly automated experimental mutagenesis procedures, disregarding \( \textit{a posteriori} \) rationalization of the experimentally observed structure–function correlations.

Among such programs, FoldX, which estimates \( \Delta \Delta G \) using a linear function of physical terms, weighted by empirically trained weight factors, performs significantly better than average in external third-party validation, with a correlation coefficient \( r \) of \( \sim 0.5 \) \( (r^2 \sim 0.3) \) across many types of structural motifs, having the most robust accuracy across all structural types. The original FoldX article reported training on a blind-test database of 625 single-point mutants in 27 simple proteins to obtain \( r \sim 0.8 \) \( (r^2 \sim 0.6) \) between predicted and experimental \( \Delta \Delta G \) of folding. Thus, despite the loss of accuracy when moving away from small, simple proteins used for parametrization, FoldX clearly recovers important aspects of the physics of protein stability.

The question then emerges of whether these methods are at all helpful in screening more complex proteins with, e.g., metal ions, several domains, and many (e.g., > 400) residues, and whether the weight factors must be reparameterized for new protein classes. The upper limit to \( r \) of \( \sim 0.86 \) \( (r^2 \sim 0.74) \) deduced from random experimental errors suggests that computational protocols could approach this limit of \( r \sim 0.8-0.9 \) \( (r^2 \sim 0.6-0.8) \) if the knowledge-based potential and the input protein structures are optimal, which is in fact obtained upon specific parametrization to known high-quality protein structures, and when using free energy perturbation/thermodynamic integration. Also, as found in this paper, there are fundamental reasons why methods such as FoldX may have nearly universal energy functions, suggesting that the stability of large, complex protein mutants far from the WT in sequence space can be predicted if the structure- and sampling-problem is solved, e.g., via Molecular Dynamics (MD) simulations.

The correlation coefficients \( r \) and, hence, the trends in stability for a series of data points are generally quite well reproduced by various computational protocols, with \( r \sim 0.9 \) \( (r^2 \sim 0.8) \) being common, \( r = 0.70-0.87 \) \( (r^2 = 0.49-0.76) \) from predictors based on the more diverse Protherm database, and \( r \sim 0.7-0.8 \) \( (r^2 \sim 0.5-0.6) \) for a recent machine learning method, and again with FoldX performing very well with \( r \sim 0.96 \) \( (r^2 = 0.92) \). Despite these good trends, average errors typically exceed 1 kcal/mol for \( \Delta \Delta G \), and qualitative predictability (more/less stable) rarely exceeds 90%, although the answer to this simple yes-or-no question is facilitated by the fact that the average, random mutation in any protein is destabilizing by \( \sim 1 \) kcal/mol.

Laccases are multicopper oxidases found in plants, bacteria, and fungi, capable of oxidizing a wide range of inorganic and aromatic substrates. In fungi, the proteins are extracellularly secreted and must thus be highly stable. Most laccases have three copper sites (T1, T2, and T3) and three copperoxin domains and possess very high reduction potentials at the mononuclear T1 site that abstracts electrons from substrates. Due to their reactivity and unusual robustness, they are increasingly used in industry, e.g., to degrade lignin for use in second-generation biofuel, for bioremediation of polluted water, or for oxidative bleaching, e.g., of dyes for use in textiles or of pulp in the paper industry. The important role of laccases motivated us to investigate whether modified protocols of FoldX can predict laccase stabilities. In this process, we investigated the sensitivity of FoldX to variable structural input and whether the FoldX energy components could be improved for laccases by means of quantitative structure property relationships (QSPR).

Two recent mutant stability data sets of the high-redox-potential fungal laccases (HRPL), TVLLiIb, a widely studied fungal laccase from \( \textit{Trametes versicolor} \) (white-rot fungi), and PM1L, have rendered such an investigation possible. The TVLLiIb data set consists mostly of single-point mutants, whereas the PM1L data set contains nine multisite mutants.

TVLLiIb and PM1L share 80% sequence identity, and both basidimycete organisms have been associated with ligninolytic processes. TVLLiIb studied in this work is the most thermostable \( (T_{opt} = 80^\circ C, p_{H_{opt}} = 2.3) \) with ABTS as substrate) of four isoforms isolated from \( \textit{Trametes versicolor} \). For an assay for temperature inactivation of ABTS-oxidation, a \( \textit{Trametes versicolor} \) isoenzyme (UniProt ID Q8TFMF1) sharing 99% identity with TVLLiIb had \( T_{opt} = 72^\circ C \).

PM1L was isolated from the fungus PM1 (CECT 2971) living in wastewater from a paper factory. It resembles other HRPLs from \( \textit{Trametes C30} \) (99% sequence identity), \( \textit{Trametes trogii} \) (97% sequence identity), and \( \textit{Coriolopsis gallica} \) (96% sequence identity). PM1L also has \( T_{opt} = 80^\circ C \) with guaiacol as a substrate, is stable at pH 3–9, and has a standard reduction potential in excess of 700 mV, making it of interest in biotechnological applications. The TVL mutants were generated by site-directed mutagenesis to improve oxidation of bulky phenolic substrates, while the PM1L mutants...
Table 1. Details of Laccase Mutant Data Sets

<table>
<thead>
<tr>
<th>protein</th>
<th>mutant name</th>
<th>specific mutations</th>
<th>thermostability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$T_{50}$ (pH = 6)</td>
</tr>
<tr>
<td>TvLIIIb</td>
<td>YL4WT</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>F162A</td>
<td>F162A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F265A</td>
<td>F265A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F332A</td>
<td>F332A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F162A/F332A</td>
<td>F162A/F332A</td>
<td></td>
</tr>
<tr>
<td>PM1L</td>
<td>6C8</td>
<td>V162A/H2O8Y/A239P/S426N/F454S/A461T</td>
<td>68.4 °C</td>
</tr>
<tr>
<td></td>
<td>6C8_S454F_REVERT</td>
<td>V162A/H2O8Y/A239P/S426N/A461T</td>
<td>72.2 °C</td>
</tr>
<tr>
<td></td>
<td>6C8_P393H</td>
<td>V162A/H2O8Y/A239P/P393H/S426N/F454S/A461T</td>
<td>62.5 °C</td>
</tr>
<tr>
<td></td>
<td>6C8_D281E</td>
<td>V162A/H2O8Y/A239P/D281E/S426N/F454S/A461T</td>
<td>66.3 °C</td>
</tr>
<tr>
<td></td>
<td>6C8_S224G</td>
<td>V162A/H2O8Y/S224G/A239P/S426N/F454S/A461T</td>
<td>66.3 °C</td>
</tr>
<tr>
<td></td>
<td>16B10</td>
<td>V162A/H2O8Y/A239P/A361T/S426N/F454S/A461T/S482L</td>
<td>71.0 °C</td>
</tr>
<tr>
<td></td>
<td>11A2</td>
<td>V162A/A239P/D281E/S426N/A461T</td>
<td>73.1 °C</td>
</tr>
<tr>
<td></td>
<td>1D11</td>
<td>H2O8Y/A239P/S426N/A461T</td>
<td>73.8 °C</td>
</tr>
<tr>
<td></td>
<td>OB-1</td>
<td>V162A/H2O8Y/S224G/A239P/D281E/S426N/A461T</td>
<td>73.1 °C</td>
</tr>
</tbody>
</table>

resulted from a strategy combining directed evolution with rational approaches to facilitate yeast expression of highly stable and active HPRL mutants. The latter approach culminated in the OB-1 mutant, which possessed a very high $k_{cat}/K_M$ for substrates such as ABTS while preserving a substantial stability at elevated $T$, diverse pH, and organic cosolvents. These promising laccases thus possess both high redox potentials and at the same time exquisite thermostability. The correlation between the experimental measures of stability and FoldX $\Delta G$ was investigated with various FoldX-protocols and MD simulations, and with a QSPR analysis of the individual energy terms comprising $\Delta G$. If reasonable correlations and acceptable errors are obtainable, it suggests a promising tool for laccase optimization, whereas the most significant energy components may be interpreted as potential physicochemical drivers of the superior stability of these laccases.

### METHODS

**Laccase Mutants.** The recent thermostability data for the WT, three single-point mutants, and one double mutant of the *Trametes versicolor* laccase TvLIIIb (UniProt ID: Q96UT7) and nine multisite mutants of the laccase PM1L (UniProt ID: Q12571) were used for benchmarking. The structure of *Trametes versicolor* laccase (PDB ID: 2HRG), which shares 97% sequence identity with PM1L. Copper ions and coordinating waters are not present in the homology model and were added manually using coordinates from the template. FoldX does not recognize carbohydrates, and the N-acetylglucosamine (NAG) moieties were removed from the structures. Omission of NAG, while reducing absolute stability, preserves high correlation between experimental and computational *relative* stability; i.e., the glycosylation effect is nearly constant and does not affect relative stabilities. This is important as it allows an understanding of the protein-specific drivers of the relative stability aside from post-translational modification.

The zero-order bond approach of the Maestro protein preparation wizard was used to fix bonds and angles involving copper to the crystal structure geometry. All copper ions were assigned a charge of +1, and the T2/T3 oxygen species were modeled as neutral water molecules, consistent with the X-ray information and preserving a realistic charge state of the protein. While exact changes in redox, protonation, and charge states are essentially impossible to model with classical force fields, for the purposes here, we were interested in the contribution to overall stability of amino acid side chains not

directly interacting with copper (at distances >8 Å from any copper site), making this study less sensitive to state-specific issues than studies aiming at understanding metal-site properties.

**Molecular Dynamics Simulations.** MD simulations were performed with Desmond 3.0 in a cubic box of TIP3P water providing a minimum layer of 15 Å of water on each side of the protein. Charge neutrality was achieved by randomly replacing waters with the required number of sodium ions. MD simulations were preceded by steepest descent minimization to a gradient of 1 kcal mol\(^{-1}\) Å\(^{-1}\) followed by the default presimulation protocol in Desmond consisting of (1) minimization with restraints on the solute; (2) unrestrained minimization; (3) Berendsen NVT simulation, \(T = 10 \text{ K}\); small time steps, restraints on heavy solute atoms; (4) Berendsen NPT simulation, \(T = 10 \text{ K}\), restraints on solute heavy atoms; (5) Berendsen NPT simulation with restraints on heavy solute atoms; and (6) unrestrained Berendsen NPT simulation.

Following the relaxation protocol, production-run NPT simulations were carried out for each system. The temperature was regulated with the Nose–Hoover chain thermostat with a relaxation time of 1.0 ps. Pressure was regulated with the Martyna–Tobias–Klein barostat with isotropic coupling and a relaxation time of 2.0 ps. The RESPA integrator was employed with bonded, near, and far time steps of 2.0 fs, 2.0 fs, and 6.0 fs, respectively. MD trajectories were saved at 100 ps intervals.

System details are provided in Table 2. Due to its larger system size and slow convergence, TvlIIib was simulated for 40 ns until the trajectory was stable, i.e., did not display divergence from the initial structure, whereas PM1L was equilibrated after 20 ns. A 9-Å cutoff was used for nonbonded interactions with the smooth-particle mesh Ewald method having a tolerance of 10\(^{-7}\) for long-range Coulomb interactions.

The root-mean-square backbone atomic positional deviation (RMSD) with respect to the initial structure (crystal or homology model) was calculated for all MD trajectory frames using VMD. The conservation of hydrogen bonds measured as presence in percentage of simulation time was calculated with the VMD “HBonds” plugin for selected donor–acceptor pairs in the PM1L simulation using as criteria a donor–acceptor distance cutoff of 3.5 Å and a donor–hydrogen–acceptor angle cutoff of 30°.

The OPLS-2005 force field was employed to describe the protein. Water and sodium ions were described with the TIP3P potential and the free-energy consistent alkali-ion OPLS potentials, respectively.

**QSPR.** QSPR for prediction of laccase thermal stability was computed using MATLAB and an in-house implementation of the NIPALS algorithm for PLS regression. Stepwise regression with forward variable selection was employed on the autoscaled matrix of the FoldX energy terms (excluding \(\Delta \Delta G\)). This resulted in successive inclusion of energy terms yielding improvements in the root-mean-square error of cross validation (RMSECV) of the PLS model. To ensure robustness, the QSPR models were restricted to employ a maximum of three FoldX \(\Delta \Delta G\) component energy terms and two latent variables, which was found to be enough to describe the stability effects. Leave-one-out cross validation (LOOCV) was employed unless otherwise noted, and \(q^2\) was evaluated for the models. A LOOCV \(q^2\) value higher than 0.5 is often regarded as a necessary (but insufficient) criterion for predictive power, but robustness can only be ascertained by external validation. The limited number of laccase mutants considered here does not allow a meaningful division into a calibration and validation set. Instead, the most promising QSPR model was subject to a second round of cross-validation based on leaving out two mutants at a time.

**RESULTS AND DISCUSSION**

**MD-Averaged Structures of TvlIIib and PM1L.** The backbone RMSD plots of the MD simulations are shown in Figure 1 (TvlIIib left, PM1L right). TvlIIib was slower to converge, and even after 30 ns, the RMSD curve still displayed a weak declination, although the RMSD fluctuations in the last part of the trajectory are confined to a relatively narrow range. The large RMSD and long simulation time for TvlIIib are due to the comparatively low resolution (2.4 Å) of the crystal structure: A substantial contribution to the RMSD comes from...
structural fluctuations in domain 3 (residues 310–499), as evident from the RMSD curves for the individual domains (Supporting Information, Figure S1). Correspondingly, the largest average crystal-structure B-factors are in domain 3 (Supporting Information, Figure S2). Also, four (Q363, R442, E460, Q482) of the five residues with missing side chain atoms in the crystal structure are located in this domain and will necessarily increase the RMSD of the all-residue MD-simulated protein against the crystal structure lacking these residues. The RMSD between the two highly homologous proteins is small (typically ∼1 Å), showing that the MD averaging produces protein structures that are comparable for our purposes.

The RMSD curve for the PM1L simulation indicates convergence after ∼10 ns with only small fluctuations at ∼1.2 Å from 10–20 ns. The fast convergence of PM1L relative to TvLIIIb in MD simulations probably reflects that the PM1L homology model was constructed from a highly homologous (97%) template with a high-resolution (1.58 Å) crystal structure.

The MD-averaged structures of TvLIIIb and PM1L are shown in Figure 2, left and right, respectively, with the mutated sites investigated in this work numbered in red colors and shown as yellow sticks and balls. Notably, as the mutations in PM1L are derived from directed evolution, there are both buried and surface sites located in various parts of the protein, suggesting that a robust prediction of stability for these mutations requires a method that performs equally well for all structural components and domains, such as has been reported for FoldX.14,24

**Structural Sensitivity of FoldX.** The correlation between $t_{1/2}$ and FoldX ΔΔG calculated at each frame of the trajectory is plotted in Figure 3 and reveals no simple relationship between MD structural convergence and predicted ΔΔG. Any particular MD snapshot is likely to produce low predictability, showing that the protein dynamics causes occupation of conformations whose apparent stabilities correlate poorly with the general, ensemble-averaged stability of the protein. FoldX is extremely sensitive to input from MD simulations because the probed phase space includes snapshots far from the statistical average that is partially preserved in, e.g., a crystal structure. Thus, the initial crystal structures are better input for stability prediction than any one MD snapshot, and using single MD snapshots for stability, property or free energy analysis should be completely avoided. Instead, it is quite meaningful to compute the average ΔΔG obtained from a number of MD snapshots (vide infra).

**Laccase Stabilities Using the Standard FoldX Protocol.** Typical use of FoldX proceeds as described in the “examples” section of the documentation, i.e., feeding a 3D structure of the enzyme into FoldX, selecting the appropriate properties, and running the protocol. The resulting ΔΔG values are then compared with experimental measurements to evaluate the accuracy of the predictions. This approach has been shown to be quite effective for predicting the stabilities of various laccase mutants, as demonstrated in Figure 3.
structure (i.e., the crystal structure of TvlIIIb or the homology model of PM1L) of the WT protein to the program, relaxing this structure ("repair" function) to remove steric clashes, and introducing new, mutated side chains with the "BuildModel" method. Additional procedures are described in the Methods section.

Application of this typical FoldX protocol to generate the TvlIIIb and PM1L mutants and compute their \( \Delta \Delta G \) relative to the WT input yielded the results depicted in Figure 4a and b, respectively. \( \Delta \Delta G \) has been plotted against experimental thermostability measures \( t_{1/2} \) and \( T_{50} \) for each set of mutants, and a least-squares line fit was evaluated. The correlation coefficient, \( r \), between \( \Delta \Delta G \) and experimental data is \(-0.90\) and \(-0.74\) for the TvlIIIb and PM1L data sets, respectively. As a larger \( \Delta \Delta G \) means less stability, corresponding to decreasing \( t_{1/2} \) and \( T_{50} \), this implies a good description of the stability trends of the overall data sets, despite substantial individual errors, fully consistent with the previous experience with FoldX and other predictors. The mainly single-site data set of TvlIIIb is more accurately reproduced by FoldX, in particular with maximum errors <1 kcal/mol and all mutants in qualitatively the right order of stability, except for the two most stable mutants. Given the fact that most of the PM1L mutants are multisite mutants differing in up to 11 sites from the WT, the FoldX correlation is markedly better than an...
Table 3. Mean Absolute, Root-Mean-Square, and Maximum Errors of FoldX Protocols Based on Non-MD-Averaged Input Structural Templates

<table>
<thead>
<tr>
<th>1-fold</th>
<th>TvlIIIb repaired</th>
<th>TvlIIIb unrepaired</th>
<th>PM1L repaired</th>
<th>PM1L unrepaired</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ΔΔG (kcal/mol)</td>
<td>t_{1/2} (min)</td>
<td>ΔΔG (kcal/mol)</td>
<td>t_{1/2} (min)</td>
</tr>
<tr>
<td>MAE</td>
<td>0.56</td>
<td>29.24</td>
<td>0.38</td>
<td>20.78</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.57</td>
<td>29.95</td>
<td>0.50</td>
<td>27.06</td>
</tr>
<tr>
<td>lMaxEl (F162A/F332A)</td>
<td>0.75</td>
<td>39.72</td>
<td>0.79 (F162A/F332A)</td>
<td>42.76 (F162A/F332A)</td>
</tr>
<tr>
<td>10-fold</td>
<td>TvlIIIb repaired</td>
<td>TvlIIIb unrepained</td>
<td>PM1L repaired</td>
<td>PM1L unrepained</td>
</tr>
<tr>
<td></td>
<td>ΔΔG (kcal/mol)</td>
<td>t_{1/2} (min)</td>
<td>ΔΔG (kcal/mol)</td>
<td>t_{1/2} (min)</td>
</tr>
<tr>
<td>MAE</td>
<td>0.58</td>
<td>31.31</td>
<td>0.35</td>
<td>18.64</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.60</td>
<td>32.09</td>
<td>0.42</td>
<td>22.31</td>
</tr>
<tr>
<td>lMaxEl (F162A/F332A)</td>
<td>0.77</td>
<td>41.61</td>
<td>0.67 (F162A/F332A)</td>
<td>35.85 (F162A/F332A)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FoldX terms (single, unrepaid runs)</th>
<th>TvlIIIb, van der Waals</th>
<th>TvlIIIb, solvation hydrophobic</th>
<th>PM1L, van der Waals</th>
<th>PM1L, solvation hydrophobic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ΔΔG (kcal/mol)</td>
<td>t_{1/2} (min)</td>
<td>ΔΔG (kcal/mol)</td>
<td>t_{1/2} (min)</td>
</tr>
<tr>
<td>MAE</td>
<td>0.49</td>
<td>28.59</td>
<td>0.74</td>
<td>28.15</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.52</td>
<td>30.09</td>
<td>0.78</td>
<td>29.59</td>
</tr>
<tr>
<td>lMaxEl (F162A/F332A)</td>
<td>0.77</td>
<td>44.53</td>
<td>1.15 (F162A/F332A)</td>
<td>43.99 (F162A/F332A)</td>
</tr>
</tbody>
</table>

"The mutants with maximum errors are given in parentheses.

average, blind, external validation, although individual errors are up to ~4 kcal/mol. The least stable mutant is correctly predicted, although the most stable mutant is not. Thus, in general, qualitative ordering of two mutants is not well modeled for the multisite mutant PM1L data set using the standard FoldX protocol, but the overall trend of the data set is well reproduced. These observations are all consistent with the general conclusions of the FoldX external validation, which suggested average errors in the ~1 kcal/mol range for single-point mutants while discouraging use for multisite mutants.19

Robustness of the Standard Protocol. Repetition of a mutation sequence in the FoldX input may be done to evaluate and improve internal statistics of FoldX. We have used this approach to obtain standard deviations for 10 repetitions for all the investigated mutations. To test the influence of local geometry optimization of the template on ΔΔG, the "repair" function was omitted in a second FoldX run on the same mutants. The omission had a minor effect for the TvlIIIb set (Figure 4e) but in fact improved r for the PM1L set from −0.74 to −0.82 (0.55 to 0.67 for r²), as seen in Figure 4d. The increased correlation was also observed for the majority of the energy terms constituting ΔΔG (see Supporting Information).

To investigate the negative effect of local optimization, we recalculated without repair for all mutations using now the average ΔΔG from 10 repeated FoldX calculations, see Figure 4e–h. For TvlIIIb, the average ΔΔG for 10 FoldX runs produced a better correlation (r = −0.94, r² = 0.88) for the unrepaired structure compared to the repaired (r = −0.89, r² = 0.79), see Figure 4g and e, respectively, whereas for PM1L, the average ΔΔG gave similar correlations with experimental data (r = −0.78, r² = 0.61) for unrepaired (Figure 4h) and repaired (Figure 4f) methods. This suggests that, although recommended for proteins inside the parametrization class, the repair function may not be useful for protein classes not used for parametrizing FoldX, probably because local template optimization indirectly affected the parametrization and also provides a bias to the template structure for proteins outside the parametrization range (vide infra).

The results in Figure 4e–h show small (<0.25 kcal/mol) standard deviations in ΔΔG for the repaired and unrepaired TvlIIIb data set. However, avoiding the repair procedure leads to larger statistical variation in calculated, repeated ΔΔG's for both data sets, and the variation is at the same time substantially larger in the PM1L data set. Thus, using repair with PM1L produces a larger deviation of ~0.5 kcal/mol (for the 6C8_P393H mutant), whereas the maximum deviation without repair is ~1.8 kcal/mol (for the 6C8 mutant). For both laccases, the standard deviation for the 10 FoldX runs approximately doubles from repaired to unrepaired, i.e., computed ΔΔG's vary much more for unrepaired structures (are less precise) most likely because the repeated calculations are better aligned to the same local conformations after optimization of the template.

However, despite the reduced precision, avoiding local optimization actually preserves or even improves correlation to experimental data (r), and error statistics (measured as distance from the regression line) are also improved, as can be seen from Table 3. The smallest maximum errors are consistently found for averages of 10 repetitions without the repair algorithm. For the PM1L data set, both the repetition protocol to improve sampling and disabling repair reduce maximum errors by ~1/2 kcal/mol. While mean absolute errors (MAE) are insignificantly affected by repetitive calculation, using the unrepaired protocol reduces MAE in all cases, by up to 0.3 kcal/mol.

In conclusion, using repetitive FoldX calculations by adding the same sequence multiple times to the FoldX input improves statistics and maximum errors, whereas using unrepaired protocols improves correlations, MAE, and maximum errors for both data sets. Both approaches are different from the standard protocol of FoldX. Thus, laccase ΔΔG prediction is improved by multiple FoldX runs using unrepaired templates. Possibly, this is true also for other, complex proteins outside the
parametrization range of FoldX: In general, whether local optimization improves or impairs depends on the quality of the input structure relative to the local optimization procedure, and local optimization may in fact produce a bias toward the initial (usually WT) template, vide infra.

**FoldX Hysteresis.** During this work, we thought that the standard FoldX protocol may be viewed as a simple form of the linear interaction free energy method of the type proposed for protein–ligand binding by Åqvist and Marelius, with sampling of the WT and mutant states combined with an empirical linear function describing the associated free energy difference based on energy components (this method can itself be viewed as an approximation to free energy perturbation schemes where intermediate coupling states are also computed and has been used for predicting protein mutant stabilities). Because of this view, we investigated whether FoldX, given that it uses a WT template as a basis for mutant generation, is biased toward the template due to relative “oversampling” (by the experimental time average) of this state. The simplest way to test this is to use a mutant sequence as the input for the generation of a structural template and compute the WT \( \Delta \Delta G \) from the mutant, i.e., performing the calculation in “reverse”.

The FoldX computations for these reverse mutations for PM1L are presented in Table 4 and are plotted against experimental \( T_{50} \) in Figure 5. On average, a reverse mutation is \( \sim 1 \) kcal/mol less stabilizing than expected from sign reversal of \( \Delta \Delta G \) for the forward mutation. This systematic error is comparable in magnitude to the maximum error observed for multiple repetitions of the same (forward) mutation from the original WT, vide supra. Performing forward mutation again, using the FoldX WT structure generated from the reverse mutation, produces a value of \( \Delta \Delta G \) that is on average \( 0.5 \) kcal/mol lower than for the initial forward mutation. The only exception is 6C8_S224G, for which \( \Delta \Delta G \) increases with \( 0.4 \) kcal/mol. This brings the predicted \( \Delta \Delta G \) of 6C8_S224G closer to 6C8_D281E, which has the same \( T_{50} \) (66.3 °C). Nevertheless, 6C8_S224G still yields the largest error (MaxE = 4.96 °C) in the prediction of \( T_{50} \), see Table 5.

This mutation was generally one of the most difficult to predict, with its stability repeatedly overestimated. To understand this, the PM1L MD trajectory was inspected around position 281 to explain why the 6C8_D281E mutant was predicted correctly. The experimental destabilization of 6C8_D281E (\( T_{50} = 66.3 ^\circ C \)) relative to 6C8 (\( T_{50} = 68.4 ^\circ C \)) can be understood by noting that the short Asp side chain provides optimal geometry for formation of the hydrogen bond Asp281-Thr190, creating an interaction between the solvent-exposed ends of two \( \beta \)-strands (Figure 6b). This hydrogen bond persists for 73% of the 20 ns MD simulation of PM1L. The additional size of Glu in the 6C8_D281E mutant impairs this interaction, and the longer Glu side chain may increase the entropy of the unfolded state and results in a larger penalty for exposure to solvent. FoldX accounts correctly for these effects, providing the correct ranking of 6C8_D281E relative to 6C8.

Altogether, these findings demonstrate that the standard repaired FoldX protocol systematically favors the initial template (typically the WT structure) over mutants by roughly 1 kcal/mol, and that this bias may be reduced by repeated

![](image)

**Figure 5.** Hysteresis in the FoldX prediction of \( \Delta \Delta G \) for PM1L mutants. \( \Delta \Delta G \) is plotted against \( T_{50} \) for (a) normal forward mutations using the PM1L homology model as WT (\( WT \rightarrow WT \)), (b) reverse mutations generating WT from mutant (\( MUT \rightarrow WT \)), and (c) forward mutations from the FoldX-generated WT (\( WT \rightarrow MUT \)).
mutation and reverse mutation. Also, relative to the repaired single PM1L calculation (Figure 4b), the hysteresis protocol results in slight improvement of the correlation between \( \Delta\Delta G \) and experimental data (\( r = -0.77 \) \( r^2 = 0.60 \) vs \( r = -0.74 \) \( r^2 = 0.55 \)) for repaired standard calculation. The errors in \( T_{50} \) prediction are smaller for the hysteresis protocol (MAE = 2.74 °C, RMSE = 3.21 °C) compared to the repaired single calculation (MAE = 2.81 °C, RMSE = 3.36 °C), see Table 5.

The fact that the majority of protein mutations are generally destabilizing is reflected in the FoldX training set, which mainly contains destabilizing mutants. This has probably affected the parametrization to cause a general bias toward the initial template during parametrization of weight factors with the repair algorithm applied. This is an additional reason for the recommendation to use our modified FoldX protocol.

Using MD-Averaged Input Structures for FoldX. We have seen that the \( \Delta\Delta G \) computations for laccases are sensitive to the FoldX protocol, including the structural input (Figure 3) and local optimization (the repair function). To investigate the influence of MD ensemble-averaging on \( \Delta\Delta G \) predictions, the last 50 frames (snapshots) from each MD simulation were used as input structures for FoldX, and \( \Delta\Delta G \) was calculated as the average of the 50 frames without use of the repair command. As seen in Figure 7a, while any particular snapshot (Figure 3) used for input produces poor results, correlation upon 50-averaging is similar to when a single crystal structure is employed (Figure 4c; \( r \sim -0.91 \), \( r^2 \sim 0.83 \)). This is consistent with MD probing dynamic changes in each frame that are averaged out in the crystal structure and, in line with the discussion above, suggests that FoldX’s local optimization impairs this time-average to some extent. Thus, local structure optimization for proteins destroys the time-average properties of the protein, including the determination of accurate \( \Delta\Delta G \). Restricted MD could possibly change this, as it has yielded good results in the CASP structure prediction competition. However, the results in Figure 7 show that one can avoid restraints to the simulations and simply calculate the average \( \Delta\Delta G \) for \( \sim 50 \) equilibrated snapshots to obtain correlations comparable to the time-averaged crystal structure.

The correlations for PM1L (Figure 7b) are similar to the single repaired run (Figure 4b) but worse than the unrepaired run (Figure 4d). A plausible explanation for the better correlation for the unrepaired run is that neither the FoldX repair function nor MD averaging can provide the time averaging afforded by the high quality crystal structure template on which the PM1L homology model was based.

For both laccases, MD compressed the \( \Delta\Delta G \) energy scale relative to non-MD FoldX runs, especially for PM1L, resulting in more realistic free energy differences. A consequence of the energy-scale compression is smaller average errors, see Table 6.

In addition, for TvLIIIb, the average error in prediction of \( t_{1/2} \) is slightly lower for the MD 50-averaging (MAE = 22.69 min, RMSE = 29.71 min) compared to the single repaired case (MAE = 29.24 min, RMSE = 29.95 min), while the maximum

---

**Table 5. Mean Absolute, Root-Mean-Square, and Maximum Errors for FoldX Hysteresis Runs on PM1L Mutants**

<table>
<thead>
<tr>
<th></th>
<th>WT→MUT (normal)</th>
<th>MUT→WT (reverse)</th>
<th>WT→MUT →WT→MUT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \Delta\Delta G ) (kcal/mol)</td>
<td>( T_{50} ) (°C)</td>
<td>( \Delta\Delta G ) (kcal/mol)</td>
</tr>
<tr>
<td>MAE</td>
<td>1.76</td>
<td>2.83</td>
<td>1.58</td>
</tr>
<tr>
<td>RMSE</td>
<td>2.09</td>
<td>3.37</td>
<td>1.89</td>
</tr>
<tr>
<td></td>
<td>MaxE</td>
<td>3.88 (S224G)</td>
<td>6.26 (S224G)</td>
</tr>
</tbody>
</table>

*The mutants with maximum errors are given in parentheses.*
error is larger for the MD-averaged case (MaxE = 49.35 min for F162A) compared to the single repaired case (MaxE = 39.72 min for F162A/F332A). The maximum error in $t_{1/2}$ prediction also increased from the repaired to unrepaired TvLIIIb structure (see above).

For PM1L, the mean errors of prediction of $T_50$ are increased by the MD averaging relative to both repaired and unrepaired single runs (Table 3), while the maximum error in $t_{1/2}$ prediction also increased from the repaired to unrepaired TvLIIIb structure (see above).

For PM1L, the mean errors of prediction of $T_50$ are increased by the MD averaging relative to both repaired and unrepaired single runs (Table 3), while the maximum error in $t_{1/2}$ prediction also increased from the repaired to unrepaired TvLIIIb structure (see above).

Quantitative Structure Property Relationships. As described in the original FoldX article, the energy terms used to compute $\Delta \Delta G$ from the 3D structure are weighted during parametrization. Since the training set excluded multicopper proteins, a substantially reduced performance for such proteins could be anticipated, as seen from the error statistics of PM1L using the standard FoldX protocol. However, the errors for TvLIIIb are comparable to optimized proteins, and the significant correlations between $\Delta \Delta G$ and experimental data for both laccases do not suggest a special problem for laccases, in particular given the possible improvements discussed above.

A complete overview of correlations between FoldX energy terms and experimental data is provided in the Supporting Information. Several energy terms correlate strongly with experimental stability for both mutants. Notably, high correlations were found for the hydrophobic desolvation term (“Solvation Hydrophobic”; $r = -0.91$ and $-0.87$ for TvLIIIb and PM1L, respectively) and for the van der Waals term ($r = \ldots$)
### Table 7. QSPR from FoldX Energy Components

<table>
<thead>
<tr>
<th>eq system</th>
<th>regression equation$^a$</th>
<th>latent variables</th>
<th>$q^2$</th>
<th>RMSECV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 PM1L repaired</td>
<td>$T_{50} = -2.31 \times [(04) \text{ van der Waals}] - 0.52 \times [(08) \text{ van der Waals clashes}] + 0.96 \times [(10)]$</td>
<td>2</td>
<td>0.84</td>
<td>1.15 °C</td>
</tr>
<tr>
<td>2 PM1L unrepair</td>
<td>$T_{50} = -1.00 \times [(07) \text{ solvation hydrophobic}] + 4.58 \times [(14) \text{ torsional clash}] + 1.99 \times [(09)]$</td>
<td>2</td>
<td>0.92</td>
<td>0.81 °C (0.88 °C)$^b$</td>
</tr>
</tbody>
</table>

$^a$The FoldX numbering of components is indicated in parentheses before each descriptor. $^b$Leave-two-out cross-validated.

−0.90 and −0.80 for TvIIIb and PM1L, respectively), as seen in Figure 8. Single FoldX energy terms were particularly successful in predicting experimental stability trends for the PM1L mutants: The hydrophobic desolvation term (i.e., the difference in aqueous desolvation and protein resolution of the residues in the mutant vs WT) produces better trends ($r = -0.87, r^2 = 0.76$) than the overall energy function ($r = -0.82, r^2 = 0.67$) and substantially reduces mean and maximum $T_{50}$ errors (MAE = 1.69 °C, RMSE = 2.11 °C, and MaxE = 3.66 °C (6C8) for hydrophobic desolvation vs MAE = 1.98 °C, RMSE = 2.42 °C, and MaxE = 4.88 °C (S224G) for the total energy).

In contrast, for TvIIIb, despite a similar correlation, the errors for $t_{1/2}$ are higher when using the single hydrophobic desolvation term (MAE = 28.15 min, RMSE = 29.59 min, MaxE = 43.99 min (F162A/F332A)) compared to $\Delta\Delta G$ (MAE = 20.78 min, RMSE = 27.06 min, MaxE = 42.76 min (F162A/F332A)). The double mutant, F162A/F332A, is always the most difficult to predict for the TvIIIb set, reflecting the general problem with prediction of multisite mutants.

Several of the FoldX energy terms are also substantially correlated among themselves. The correlation $r$ between hydrophobic solvation and van der Waals terms is 0.97 and 0.93 for unrepaired and repaired structures, respectively, explainable by the fact that most of the studied site mutations preserve chemical properties (charge, polarity) as the WT (see Table 1). The electrostatic term is less important in the studied cases, since for TvIIIb, all mutations are F to A, i.e., purely hydrophobic, and most of the mutations of the PM1L data set are either polar to polar, unipolar to unipolar, or charged to charged. Thus, the energy terms performing well for the present mutations are not likely to be optimal for mutations that markedly change the properties of the site; such mutations tend to have low probability, low evolution rate, and commonly lead to unstable or nonfunctional protein.

Since one term can describe most of the correlation and reduce errors, the FoldX energy function may be compressed for laccases. This is consistent with more advanced sampling in linear interaction free energy methods where only a few terms, usually electrostatic and van der Waals, suffice,$^{66}$ and with the fact that from a physical point of view, there are only two nonbonded energy terms present in standard molecular modeling force fields, i.e., an electrostatics and a van der Waals term.

To explore these issues further, QSPR was established to investigate the room for improvement by a different weighting of energy terms. Due to its small size, the TvIIIb data set was not considered in QSPR analysis. Two descriptor matrices for the PM1L mutants were employed, comprised by the FoldX energy terms calculated using the repaired and unrepaired homology model, respectively.

The least stable mutant, 6C8_P393H ($T_{50} = 62.5 ^\circ C$) was an outlier in the first generation of QSPR models, and was therefore excluded from subsequent models. While not an outlier in $\Delta\Delta G$ prediction, although the error in $\Delta\Delta G$ is often large, the QSPR models are restricted to employ a maximum of three FoldX $\Delta\Delta G$ component energy terms, preventing the inclusion of noisy descriptors that might otherwise have reproduced the particular instability of the 6C8_P393H mutant by overfitting. Inspection of the MD-averaged PM1L structure (Figure 2 right and Figure 6c) shows that P393 is found at the proximal end of a loop comprised by residues 383−394 partially covering the T1 site. The P393H mutation likely impairs the backbone conformation of this loop and increases its flexibility, consequently lowering the thermostability of the mutant while increasing its activity due to increased substrate access (smaller $K_M$). This mechanism of destabilization is unlikely to be represented by a few FoldX energy terms such as in the QSPR models below.

The parameters for the QSPR models generated for the eight remaining PM1L mutants are shown in Table 7, and the predicted versus measured plots for the models are shown in Figure 9. Both the QSPR models based on the repaired and the unrepaired structure consist of two latent variables and three energy terms. Both models are of high quality, although $q^2$ for the unrepaired model ($q^2 = 0.92$) is higher than for the repaired model ($q^2 = 0.84$). This is consistent with our recommendation of the unrepaired approach for laccases based on overall FoldX calculations.

As suspected from the individual terms (Figure 8), the most important terms included into each QSPR model were the van der Waals term and the hydrophobic solvation term for the repaired and unrepaired protocols, respectively. In both QSPR models, the second and third terms are a steric and an entropic term. For the repaired protocol, van der Waals clashes, representing inter-residue close contacts, are important, whereas the steric term for the unrepaired protocol is the torsional clash, representing intraresidue conformational strain. These terms are uncorrelated and most likely differ because the repair protocol relaxes some local strain. Significant correlation ($r = 0.85, r^2 = 0.72$) exists between the main-chain and side-chain entropic terms (terms 10 and 9 in the FoldX function) for the repaired and unrepaired model. Thus, the QSPR models for both protocols reflect the same physics, with exception of the descriptor encoding steric clashes due to the difference in relaxation repair, which is not surprising as the repair function minimizes steric clashes.

The best QSPR model was obtained from an unrepaired protocol (Figure 9b): The RMSE for $T_{50}$ predictions based on regression to FoldX $\Delta\Delta G$ is 1.98 °C, while the RMSECV of the QSPR is 0.81 °C for leave-one-out cross validation. Although the exclusion of the 6C8_P393H mutant from the QSPR model excludes a strict comparison, we note that the QSPR model is based on only three energy terms (Table 7, eq 2) and is robust under cross-validation when leaving two samples out (Figure 9c) with a leave-two-out-RMSECV of 0.88 °C. Thus, the physical terms defining this QSPR model are likely to also define the experimentally observed stability of the laccases. This insight should be of value in understanding contributions to
laccase stability and may be useful in predicting more stable multisite mutants of fungal laccases relatively far away from the WT in sequence space.

In experiments, the F454S mutation caused significant thermal destabilization while at the same time substantially increasing enzyme activity. This stability-activity trade-off was explained by noting that Ser in the F454S mutant enables a new hydrogen bond, probably exerting a pull on the helical segment, where one of the T1 copper coordinating histidines (His555) is located. The elongated His—Cu bond distance reduces electron density and thus increases the electronegativity (i.e., redox potential) of the T1 copper but reduces enzyme stability. Experimental reversal of F454S yielded the 6C8_S454F_REVERT mutant with fully restored thermal stability. In the present work, FoldX correctly produces a substantially smaller ΔΔG for the 6C8_S454F_REVERT mutant relative to 6C8. Since FoldX does not explicitly model the ligand—metal interaction, the qualitatively correct ΔΔG is likely due to significant stabilizing contributions from Phe to the FoldX terms “Solvation Hydrophobic” and “van der Waals” as seen from the scatter plots of these descriptors versus experimental stability in Figure 8.

The 16B10 mutant was another result of efforts to improve the thermostability of PM1 laccases.16B10 differs from 6C8 by two mutations, A361T and S482L, and is substantially more thermostable (T50 = 71.0 °C) than 6C8 (T50 = 68.4 °C) with ~1/3 activity. This is yet another example of the stability—activity trade-off in laccases that would be of substantial interest to understand in molecular terms.

Stabilization mechanisms by A361T and S482L differ as one is a hydrophobic to polar, the other a polar to hydrophobic mutation: Ala361 is exposed in the loop composed by residues Thr361—Val370, and the introduced Thr361 hydrogen bonds to Val370 C==O and the Ser372 side chain, stabilizing this loop. In contrast, Ser482 is located at the beginning of the C-terminal α-helix. A hydrogen bond between Ser482 and the backbone C==O of Pro478 is observed in the PM1L MD trajectory (Supporting Information, Figure S3A). Disruption of the i—fourth (or i + fourth) backbone hydrogen bond in α-helices is common for serine and is associated with increased flexibility and plausibly destabilization. The S482L mutation is likely favorable because Leu does not form α-disruptive interactions. Also, the hydrophobic Leu side chain may have more favorable hydrophobic interactions,33 which upon inspection of the MD-averaged structures are the nearby hydrophobic residues Ala477 and Leu482. The additional Leu is unfavorable in the unfolded state, where hydrophobic side chains incur a penalty favoring folding.

FoldX in its standard version assigns similar ΔΔG within the uncertainty of the method for 6C8 and 16B10 except in the unrepaired runs, where the correlation agrees well with measured T50 (Figure 4, data points at 68 and 71 °C). Similar stability is also obtained by direct comparison of the mutants using the 6C8 structure as FoldX input to generate 16B10 (the five-repetition average ΔΔG is 0.1 kcal/mol for the double mutation). Thus, despite having the correct trends of the two mutants in the overall data set, FoldX cannot reproduce a significant stabilizing effect of the double-mutation from 6C8 to 16B10.

To understand the drivers of the experimentally observed difference, the two single-step mutations were instead studied individually in FoldX as 6C8_A361T and 6C8_S482L from 6C8. The average ΔΔG for the 6C8_A361T mutant was 2.0 kcal/mol, implying that FoldX does not predict the stabilization mode of Thr361 discussed above. This is explained by the FoldX-generated 6C8_A361T structure, where hydrogen bonding with Ser372 is disallowed by its orientation (Supporting Information, Figure S3B). However, the average ΔΔG (−1.9 kcal/mol) for S482L indicates that FoldX captures the stabilizing effects discussed above for this mutation. When these mutations are done together, the net result is nearly neutral in FoldX due to the poor sampling of the hydrogen bond introduced by A361T. Substantial stabilizing contributions from “Solvation Hydrophobic” and “van der Waals” suggest that the principal stabilizing effect is the increased hydrophobicity of the Leu side chain more than the disruption itself, which is why FoldX succeeds for this mutation (full energy components are given in Supporting Information, Tables S3—S5). When run in reverse, the above mutations showed no artificial stabilization of the FoldX-generated 6C8 input structure. In summary, the precise ranking of 6C8 or 16B10 depends on the ability to sample well the hydrogen.
bond introduced by A361T, whereas S482L is well-modeled. While giving some insight into the drivers of the stability of the two mutations, a more accurate quantification might be obtained by future optimization of the hydrogen-bond sampling in FoldX.

The PM1L mutant set, due to its size and considerable variation in both stability and activity, may be useful for analyzing the stability/activity trade-off in laccases. However, it is not meaningful to evaluate the correlation directly, since the activity is affected by the combined effect of mutations in the α-factor prepro-leader as well as the mature protein, whereas stability pertains to the mature expressed protein only. For example, 7H2 and 6C8 differ only by a single mutation in the prepro-leader and have total activity increases (TAI) of 24 290 and 43 720, respectively. For mutations with the same prepro-sequence, the following was found: 6C8 (TAI = 43 720), 6C8_S454G (TAI = 21 860), 6C8_P393H (TAI = 65 880), 6C8_D281E (TAI = 56 836), 6C8_S224G (TAI = 56 836), and OB-1 (TAI = 34 000). As shown in Figure 10a, TAI is nearly perfectly anticorrelated \( r = -0.94, r^2 = 0.88 \) with \( T_{50} \) for these mutants. Interestingly, the hydrophobic solvation term correlates significantly with TAI (Figure 10b). These types of relations may be useful in circumventing the stability-function trade-off in protein engineering, in this case specifically for laccases.

**CONCLUSIONS**

In this work, we have examined if it is possible to reproduce and rationalize the stability of industrially important high-redox potential fungal laccases TvLIIIb and PM1L. A standard FoldX protocol using a single, repaired crystal structure (TvLIIIb) or homology model (PM1L) as structural input gave correlation coefficients, \( r \), between \( \Delta \Delta G \) and experimental data of \( -0.90 \) and \( -0.74 \) (0.81 and 0.55 for \( r^2 \)) for the TvLIIIb and PM1L data sets, respectively. These correlations imply that standard FoldX use provides a fair description of the stability trends of the overall data sets, although individual errors are in some cases substantial (up to \( \sim 4 \) kcal/mol). As expected, the multistart set PM1L was more difficult to predict than the predominantly single mutant TvLIIIb set.

Local optimization (the “repair” function) was omitted in a second FoldX run, surprisingly yielding a better correlation \( r = -0.82, r^2 = 0.67 \) with experimental data for the PM1L set and a minor improvement for the TvLIIIb set \( r = -0.92, r^2 = 0.85 \) relative to repaired runs. Furthermore, the mean absolute and root-mean-square errors of prediction of \( t_{1/2} \) and \( T_{50} \) were lowered relative to the repaired runs. Similar improvements were observed for the energy components comprising \( \Delta \Delta G \).

Robustness of the predictions was probed by 10-fold repetition of the FoldX calculations. This resulted in a 10-average of \( r = -0.94 \) (0.88) for the unrepaired protocol compared to \( r = -0.89 \) (0.79) for the repaired protocol for TvLIIIb. For PM1L, correlations were similar \( r = -0.78, r^2 = 0.61 \) for both methods. Omitting the repair function approximately doubled the standard deviation (i.e., decreased the precision) for both data sets but reduced mean errors (i.e., improved accuracy) in all cases by up to 0.3 kcal/mol and universally produced smaller maximum errors. Thus, the FoldX repair function is less applicable to proteins outside the parametrization range of FoldX.

A bias toward the input template was identified in a series of WT to mutant (“forward”) and mutant to WT (“reverse”) FoldX calculations for the PM1L set, with reverse mutations on average \( \sim 1 \) kcal/mol less stabilizing than from sign reversal of \( \Delta \Delta G \) for the forward mutations. The WT bias may originate from the dominance of destabilizing mutations in the FoldX training set. A second round of forward mutations using FoldX generated WT structures as input reduced bias by \( \sim 0.5 \) kcal/mol and reduced average errors of \( T_{50} \) (MAE by 0.07 °C, RMSE by 0.15 °C) relative to the standard protocol. Thus, successive forward and reverse mutations (or the average of one forward and one backward calculation) may reduce FoldX’s template bias to enhance accuracy outside the parametrization range of FoldX.

A strong dependence of FoldX predictions on protein conformation was demonstrated using snapshots from the equilibrated part of the MD trajectories: Protein dynamics produces conformations whose apparent stabilities correlate poorly with the ensemble-averaged stability of the protein (i.e., \( > 1 \) kcal/mol). Thus, single MD snapshots yielded poor correlations with \( \Delta \Delta G \), whereas the average \( \Delta \Delta G \) from 50 individual FoldX calculations on distinct snapshots was comparable to results from the crystal structure: 50 MD snapshots restore correlation \( r = 0.7–0.9, r^2 = 0.5–0.8 \) and provide a root-mean-square accuracy of \( \sim 1.2 \) kcal/mol for \( \Delta \Delta G \) or 3.5 °C for \( T_{50} \) for the PM1L data (Table 6), suggesting that the time average of the crystal structure is mimicked. In lieu of a crystal structure, a high quality homology model or an ensemble average over ca. 50 equilibrated MD snapshots will be uninformative.
We found that several FoldX components (notably “Solvation Hydrophobic” and “van der Waals”) correlate strongly with experimental laccase stabilities, lending promise of a simplified predictor of laccase stability. Quantitative structure property relationships gave excellent predictions (\(q^2 = 0.92\), RMSECV = 0.81 °C) with a small subset of FoldX energy terms, demonstrating that FoldX effortlessly can be recalibrated to predict stabilities for new classes of proteins. No attempts were made to model the potentially extensive glycosylation found in the native laccases, yet the correlations between FoldX predictions and experimental stabilities were comparable to or better than for random proteins in previous benchmarks. This suggests that glycosylation may shift a common stability offset for all the mutants but has a negligible effect on their relative stabilities.

**ASSOCIATED CONTENT**

Supporting Information
Detailed RMSD analysis for the TvLIIIb MD simulation: Figure S1. Domain based RMSD curves for 40 ns TvLIIIb MD simulations. Figure S2. B-factors averaged over backbone atoms for residues in molecule A in the crystal structure (PDB-ID: 1KYA) of *Trametes versicolor* laccase. Structural exploration of selected mutations: Figure S3. Snapshot from the PM1L MD simulation showing a transient hydrogen bond between the side chain of Ser482 and the backbone oxygen of Pro478. FoldX-generated structure for the 16B10 mutant, showing the hydrogen bond between the Thr361 side chain and the Val370 backbone oxygen. It is also evident how the rotation of the Ser372 side chain precludes the formation of a hydrogen bond to Thr361. Correlations between FoldX energy terms and experimental thermostability data for mutants of the laccases TvLIIIb and PM1L: Table S1. Correlations between FoldX energy terms and the experimental stability measure (\(t_{1/2}\)) for TvLIIIb mutants. Table S2. Correlations between FoldX energy terms and the experimental stability measure (\(T_{50}\)) for PM1L mutants. Tables S3–S5: Energy components from FoldX for analyzing 16B10. This material is available free of charge via the Internet at http://pubs.acs.org

**AUTHOR INFORMATION**

Corresponding Author
*E-mail: kpj@kemi.dtu.dk.

Notes
The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

The project has been made possible by funding from the Danish Council for Independent Research: Technology and Production Sciences (FTP), Case Number 10-082488.

**REFERENCES**

(28) Dwivedi, U. N.; Singh, P.; Pandey, V. P.; Kumar, A. Structure—
function relationship among bacterial, fungal and plant laccases. J. Mol.
(29) Quintanar, I.; Stoj, C.; Taylor, A. B.; Hart, P. J.; Kosman, D. J.;
Solomon, E. I. Shall we dance? How a mult copper oxidase chooses its
Laccase Engineering Database: A classification and analysis system for
laccases and related multicopper oxidases. Database 2011, bar016.
(31) Rodríguez, C. S.; Herrera, L. J. T. Industrial and biotechnolo-
(32) Mayer, A. M.; Staples, R. C. Laccase: new functions for an old
(33) Stoi lova, I.; Krastanov, A.; Stanchev, V. Properties of crude
laccase from Trametes versicolor produced by solid-substrate
(34) Jolivalt, C.; Madzak, C.; Bault, A.; Caminade, E.; Malosse, C.;
Mougin, C. Expression of laccase IIIb from the white-rot fungus
Trametes versicolor in the yeast Yarrowia lipolytica for environmental
(35) Coll, P. M.; Tabernero, C.; Santamaria, R.; Pérez, P. Characterization and structural analysis of the laccase I gene from
the newly isolated ligninolytic basidiomycete PM1 (CECT 2971).
(36) Galli, C.; Gentili, P.; Jolivalt, C.; Madzak, C.; Vadali, R. How is
the reactivity of laccase affected by single-point mutations? Engineer-
ing laccase for improved activity towards sterically demanding
Camarero, S. Alcalde, M. Laboratory evolution of high-redox potential
(38) Koschorreck, K.; Richter, S. M.; Swierzcek, A.; Beifuss, U.;
(39) Coll, P. M.; Fernández-Abalos, J. M.; Villanueva, J. R.;
Santamaria, R.; Pérez, P. Purification and characterization of a
phenoloxidase (laccase) from the lignin-degrading basidiomycete
(40) Bertrand, T.; Jolivalt, C.; Briozzo, P.; Caminade, E.; Joly, N.;
Madzak, C.; Mougin, C. Crystal structure of a four-copper laccase
complexed with an aryamine: Insights into substrate recognition and
(41) Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J. F.; Honig,
B.; Shaw, D. E.; Friesner, R. A. A hierarchical approach to all-atom
367.
(42) Schwede, T.; Kopp, J.; Güex, N.; Peitsch, M. C. SWISS-
MODEL: An automated protein homology-modeling server. Nucleic
(43) Kiefte, F.; Arnold, K.; Künzli, M.; Bordoli, L.; Schwede, T. The
SWISS-MODEL repository and associated resources. Nucleic Acids Res.
(44) Desmond Molecular Dynamics System, version 3.0; D. E. Shaw
(45) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.;
Klein, M. L. Comparison of simple potential functions for
(46) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.;
DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an
(48) Hoover, W. G. Canonical dynamics: Equilibrium phase-space
(50) Tuckerman, M.; Berne, B. J.; Martyna, G. J. Reversible multiple
(51) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular
(52) Jorgensen, W. L.; Turro-Rives, J. The OPLS force field for
proteins. Energy minimizations for crystals of cyclic peptides and
(53) Banks, J. L.; Beard, H. S.; Can, Y.; Cho, A. E.; Dannm, W.; Farid,
Philipp, D. M.; Repasky, M. P.; Zhang, L. Y.; Berne, B. J.; Friesner, R.
A.; Gallicchio, E.; Levy, R. M. Integrated Modeling Program, Applied
(54) Jensen, K. P.; Jorgensen, W. L. Halide, ammonium, and alkali
metal ion parameters for modeling aqueous solutions. J. Chem. Theory
(55) MATLAB, version 7.10.0; The MathWorks Inc.: Natick, MA,
2010.
(56) Geladi, P.; Kowalski, B. R. Partial least squares regression (PLS):
(57) Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis.
(58) Golbraikh, A.; Troshpa, A. Beware of q2! J. Mol. Graphics
(59) Troshpa, A. Best practices for QSAR model development,
(60) Åqvist, J.; Marelus, J. The linear interaction energy method for
predicting ligand binding free energies. Comb. Chem. High Throughput
Screening 2001, 4, 613–626.
energy method for the prediction of protein stability changes upon
(62) Flehol, J. A.; Fried, G.; Berendsen, H. J. C. Completion and
refinement of 3-D homology models with restricted molecular
dynamics: application to targets 47, 58, and 111 in the CASP
modeling competition and posterior analysis. Proteins: Struct., Funct.,
(63) García-Ruiz, E.; Mate, D.; Ballesteros, A.; Martínez, A. Alcalde,
M. Evolving thermobility in mutant libraries of ligninolytic
(64) Hall, S. E.; Roberts, K.; Vaideli, N. Position of helical kinks in
membrane protein crystal structures and the accuracy of computa-