Pushing the limits of automatic computational protein design: Design, expression, and characterization of a large synthetic protein based on a fungal laccase

ARTICLE *in* SYSTEMS AND SYNTHETIC BIOLOGY · JUNE 2011

DOI: 10.1007/s11693-011-9080-9 · Source: PubMed

6 AUTHORS, INCLUDING:

Geza Szilvay
VTT Technical Research Centre of Finland
29 PUBLICATIONS **951** CITATIONS
SEE PROFILE

María Suarez-Diez
Wageningen University
38 PUBLICATIONS **225** CITATIONS
SEE PROFILE

Alfonso Jaramillo
The University of Warwick
86 PUBLICATIONS **785** CITATIONS
SEE PROFILE

Scott Banta
Columbia University
102 PUBLICATIONS **1,082** CITATIONS
SEE PROFILE

Syst Synth Biol (2011) 5:45–58
DOI 10.1007/s11693-011-9080-9

METHODOLOGY ARTICLE

Pushing the limits of automatic computational protein design: design, expression, and characterization of a large synthetic protein based on a fungal laccase scaffold

Doris J. Glykys · Géza R. Szilvay · Pablo Tortosa · María Suárez Diez · Alfonso Jaramillo · Scott Banta

Received: 18 May 2010 / Revised: 12 November 2010 / Accepted: 19 February 2011 / Published online: 20 March 2011

Abstract The de novo engineering of new proteins will allow the design of complex systems in synthetic biology. But the design of large proteins is very challenging due to the large combinatorial sequence space to be explored and the lack of a suitable selection system to guide the evolution and optimization. One way to approach this challenge is to use computational design methods based on the current crystallographic data and on molecular mechanics. We have used a laccase protein fold as a scaffold to design a new protein sequence that would adopt a 3D conformation in solution similar to a wild-type protein, the *Trametes versicolor* (TvL) fungal laccase. Laccases are multi-copper oxidases that find utility in a variety of industrial applications. The laccases with highest activity and redox potential are generally secreted fungal glycoproteins. Prokaryotic laccases have been identified with some desirable features, but they often exhibit low redox potentials. The designed sequence (DLac) shares a 50% sequence identity to the original TvL protein. The new DLac gene was overexpressed in *E. coli* and the majority of the protein was found in inclusion bodies. Both soluble protein and refolded insoluble protein were purified, and their identity was verified by mass spectrometry. Neither protein exhibited the characteristic T1 copper absorbance, neither bound copper by atomic absorption, and neither was active using a variety of laccase substrates over a range of pH values. Circular dichroism spectroscopy studies suggest that the DLac protein adopts a molten globule structure that is similar to the denatured and refolded native fungal TvL protein, which is significantly different from the natively secreted fungal protein. Taken together, these results indicate that the computationally designed DLac expressed in *E. coli* is unable to utilize the same folding pathway that is used in the expression of the parent TvL protein or the prokaryotic laccases. This sequence can be used going forward to help elucidate the sequence requirements needed for prokaryotic multi-copper oxidase expression.

Keywords Laccase · Multi-copper oxidase · Computational protein design · Protein folding · Molten globule

Introduction

One of the most difficult challenges in synthetic biology is the de novo design of proteins with targeted structures and functionalities. This requires the production of a novel amino acid sequence, preventing in most cases the use of experimental methods such as directed evolution to find such sequence. This is due to the fact that the combinatorial space is often too large and efficient evolution strategies require some fitness to start with. We have explored the possibility of using computational methods for this purpose instead. This enables the exploration of a large combinatorial space thanks to smart pair-wise algorithms that allow more efficient searching of the sequence space and this is facilitate by the use of distributed computing. In this

Electronic supplementary material The online version of this article (doi:10.1007/s11693-011-9080-9) contains supplementary material, which is available to authorized users.

D. J. Glykys · G. R. Szilvay · S. Banta (✉)
Department of Chemical Engineering, Columbia University in the City of New York, 820 S. W. Mudd, MC 4721, 500 W 120th St, New York, NY 10027, USA
e-mail: sbanta@columbia.edu

P. Tortosa · M. Suárez Diez · A. Jaramillo
ISSB, Genopole-Université d'Évry Val d'Essonne-CNRS UPS3201, Batiment Geneavenir 6. 5, rue Henri Desbruères., 91030 Evry Cedex, France

manuscript we report this approach with a much larger system than has previously been reported in the literature, the laccase fold.

Laccases are multi-copper oxidase enzymes that catalyze the oxidation of a wide range of substrates including phenolics, polyphenolics, aromatic amines, and other compounds with the concomitant reduction of dioxygen to water (Kunamneni et al. 2008; Mayer and Staples 2002; Nakamura and Go 2005; Riva 2006; Sakurai and Kataoka 2007; Solomon et al. 1996). They have many important industrial applications including: textile and fabric bleaching, lignin degradation for paper production, pollution detoxification, wine clarification, organic chemical synthesis, biosensing and the cathodic reaction in enzymatic biofuel cells.

The most commonly used laccase enzymes for biofuel cells and other industrial applications are the lignolytic glycoproteins secreted by white-rot fungi such as *Trametes versicolor* (Piontek et al. 2002). The *Trametes versicolor* laccase (TvL) is a glycosylated monomer with 499 amino acids and a molecular mass of $\sim$70 kDa. Like other laccases, the protein contains two copper-containing active centers. The first contains a T1, or blue copper center, where two histidines and a cysteine serve as ligands for the copper and this active site is where the oxidation of the substrate takes place. The second copper center is a trinuclear T2/T3 site where eight histidine side chains serve as ligands for a T2 copper and a pair of T3 coppers and in this site the reduction of dioxygen to water occurs. Interestingly, the T2/T3 site is buried deeply in the protein, and is accessible via a tunnel that opens on both sides of the enzyme. Besides the cysteine side chain that coordinates the T1 copper, there are four other cysteines in the sequence that form two disulfide bonds. The enzyme follows the bi-bi ping-pong kinetic mechanism, and it is most active at acidic pH values (Xu 1997; Xu et al. 1996).

TvL and other fungal laccase enzymes tend to have high redox potentials which allow them to oxidize a wide range of substrates, and in a biofuel cell, the high redox potentials are desirable as they lead to high operating voltages (Barton et al. 2004; Glykys and Banta 2009; Shleev et al. 2005). Some of the fungal enzymes have been expressed recombinantly in yeast, and there are a few reports where protein engineering has been performed to attempt to improve their expression, redox potential, and activity (Gelo-Pujic et al. 1999; Madzak et al. 2006; Rodgers et al. 2010; Xu et al. 1998, 1999). But, in published applications involving laccases, the native enzymes from the fungal hosts are most often utilized.

In addition to the fungal enzymes, there are many prokaryotic laccase and laccase-like proteins, and the native functions of these proteins are less well-understood (Alexandre and Zhulin 2000; Claus 2003). There has been a good deal of success in cloning and expressing these enzymes recombinantly in *E. coli*, which means that copper-containing active sites can be formed and properly folded in prokaryotic hosts. These enzymes could have significant advantages over the fungal enzymes as they could be readily produced economically on a large scale, they are not glycosylated, and they often are active over a range of pH values including neutral pH. In addition, bacterial expression can simplify protein engineering techniques such as directed evolution (Brissos et al. 2009; Bulter et al. 2003; Festa et al. 2008; Gupta and Farinas 2009).

Unfortunately, one of the hallmarks of the prokaryotic laccases is that they tend to exhibit low redox potentials (Endo et al. 2003; Li et al. 2007; Martins et al. 2002). This limits their use in industrial applications, and can significantly reduce their utility in the creation of enzymatic biofuel cells as this leads to a reduction in voltage and therefore, power. For example, a bacterially-derived laccase enzyme (SLAC from *Streptomyces coelicolor*) (Machczynski et al. 2004; Skalova et al. 2009) expressed in *E. coli* has recently been incorporated into an osmium-mediated biofuel cell cathode (Gallaway et al. 2008; Wheeldon et al. 2008). Under acidic conditions its performance was reduced compared to the same cathode made with the TvL enzyme. But, at neutral pH, the highest current density for an enzyme mediated cathode was reported (Gallaway et al. 2008). In order to address this limitation of the prokaryotic laccases, other researchers are working to use protein engineering to attempt to increase the redox potential of the prokaryotic laccase-like proteins to improve their performance (Brissos et al. 2009).

It is likely that other researchers have tried to express the TvL glycoprotein in *E. coli*, but there has yet to be a report of the successful expression of this enzyme in a prokaryotic host (Kunamneni et al. 2008). There may be many reasons for this lack of success, including problems with stability, folding, or solubility of the protein in the prokaryotic host. Such expression requires, among other things, to ensure a proper stability, folding and solubility. One way to achieve such specifications is by using computational design, where automatic methods are used to explore the sequence space to find the optimal solution given an objective function. An objective function ensuring stability and solubility is given by the folding free energy. This requires using a structure template for the folded structure, which could be accomplished by choosing a tertiary structure with the appropriate fold. By picking an existing structure we increase the probabilities of maintaining the folding pathways, although there is currently no general methodology to account for it. However, since the fungal laccase structures are very similar to the prokaryotic laccases, we reasoned that a protein with an identical

structure to the TvL protein should be expressible in *E. coli*., in an active form, if properly designed. One way to explore such specifications is by using computational design, where automatic methods are used to explore the sequence space to find the optimal solution given an objective function. An objective function ensuring stability and solubility is given by the folding free energy. This requires using a structure template for the folded structure, which could be accomplished by choosing a tertiary structure with the appropriate fold. By picking an existing structure we increase the probabilities of maintaining the folding pathways, although there is currently no general methodology to account for this.

Here we report the use of a computational approach to attempt to redesign the TvL sequence for increased stability and for expression in a prokaryotic host. The computational approach begins with the TvL protein backbone structure, and new amino acid side chains are identified with the goal of stabilizing this scaffold, thus forming a new protein with the same 3-D structure as the glycosylated TvL. This methodology has been demonstrated through the redesign of 45 proteins which were successfully compared with their corresponding natural sequences.(Jaramillo et al. 2002) The new TvL-based sequence constitutes, to our knowledge, the largest computational design projects ever attempted with a sequence space of 10^{341} possible states. This size is too large to only use the current pair-wise optimization procedures and it required the use of extensive heuristics.

The newly designed laccase (DLac) gene was synthesized and the gene product was overexpressed in *E. coli* and the majority of the protein was found in inclusion bodies (IBs). Both solubly expressed and refolded proteins were purified and characterized, and the protein appears to be trapped in a molten globule structure that does not bind copper and is inactive under the experimental conditions tested. This molten globule structure seems to be very similar to what is observed when native TvL is chemically denatured and refolded, which suggests that proper folding of the TvL (and DLac) backbone requires a folding pathway that is not enabled by the sequence, or that may not be available in *E. coli*. This new gene will be a valuable starting point for unraveling the sequence requirements for successful laccase folding in prokaryotic hosts.

Materials and methods

Computational design considerations

Computational protein design was used to create a new primary sequence that would adopt the same structure as the secreted TvL protein. Ideally this would result in a protein product with the kinetic activity and redox potential of the fungal parent enzyme with the ability to be expressed in a highly stable and active form in a prokaryotic host. TvL was chosen as the starting structure due to its high redox potential (+780 mV vs. NHE) and the availability of a 1.90 Å resolution crystal structure (PDB code 1GYC) (Piontek et al. 2002). The structure of the oxidized form of the laccase contains the full complement of coppers needed for its catalytic activity.

Copper atoms are critical for the catalytic function of the enzyme, and the native TvL enzyme has a tunnel through the protein to allow the transport of oxygen and water to and from the trinuclear T2/T3 site active site. Since these features are crucial for the function of the enzyme, the amino acids at the 145 positions within a 12 Å radius of the four copper atoms were fixed to have the same amino acid identity as in the parent sequence. Since proline and cysteine residues can have large impacts on the overall structure of the protein, those residues were also left unavailable for mutation. To add further constraints, the 51 amino acids within 8 Å from the cooper atoms and those that formed tight interactions on the surface of the structure (37 positions) were left unmutated and were also forced to have the same structural conformation as observed in the crystal structure. The 8 Å radius was chosen since the Charmm energy force field has its parameters optimized to use a cutoff of 7 Å for the non-bonded interactions and a 1.4 Å solvation probe. Therefore, 8 Å is the minimum cutoff around the copper atoms that would not change the active site energetic. Outside of these overlapping regions, all of the non-proline and non-cysteine residues were allowed to be substituted to new amino acids with variable conformations. From the original 499 amino acids in the crystal structure, there were 57 positions where the identity was retained but conformational changes were allowed, and a total of 339 positions (designed positions) where mutations were allowed which leads to a combinatorial space of $20^{339} \approx 10^{341}$ possible sequences for the protein design (Fig. 1).

Computational protein design procedure

The computational protein design software, DESIGNER, has previously been described and validated (Jaramillo et al. 2002; Ogata et al. 2003; Wernisch et al. 2000). The DESIGNER algorithm approaches computational protein design through the inverse folding problem. It begins with a high resolution protein target structure, and the available sequence space is searched to find new amino acid combinations that have optimized intramolecular interactions and will thus cause the protein to adopt the same fold as the desired targeted structure. For each analyzed sequence,

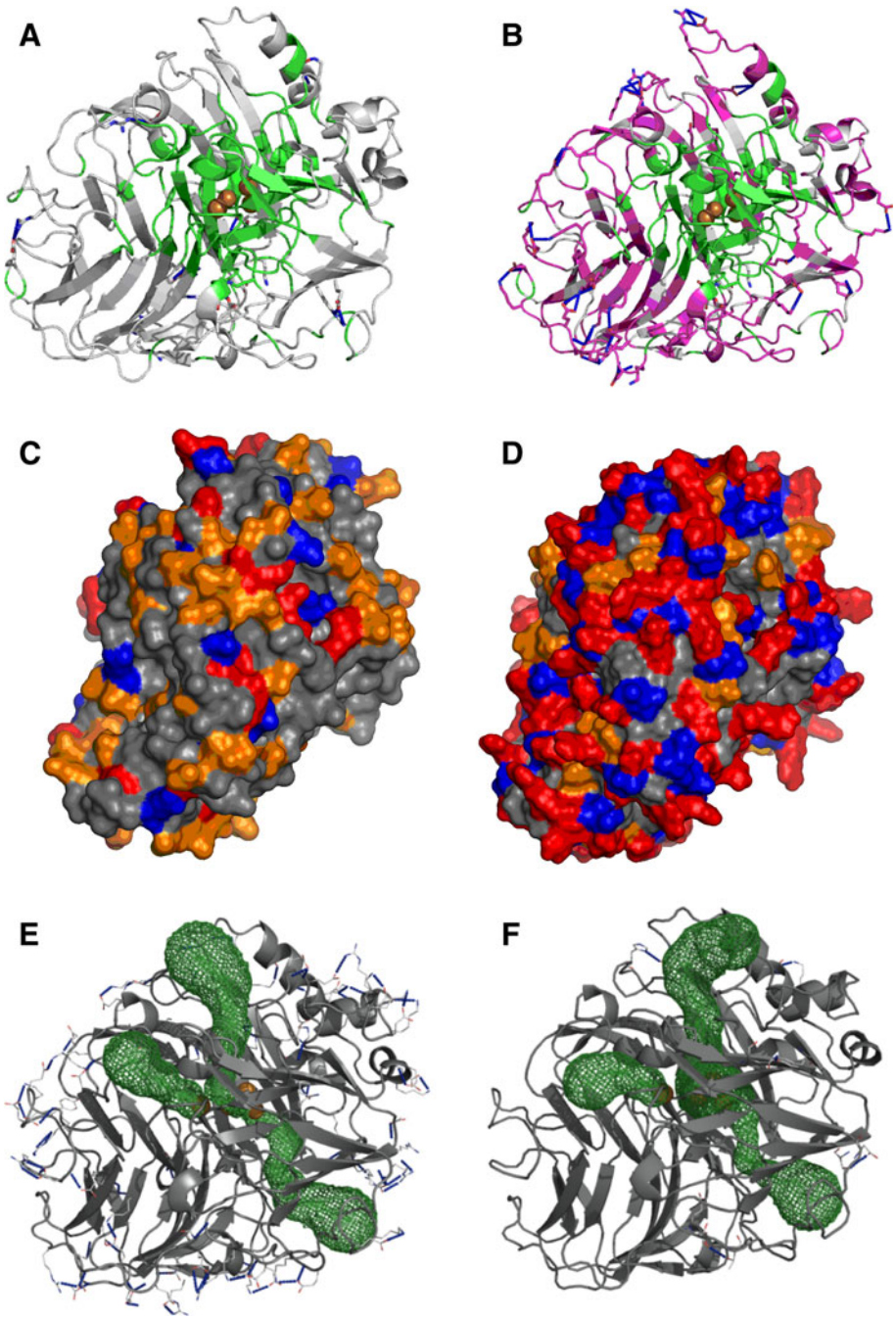

Fig. 1 Structures of the TvL and DLac proteins. **A** 3-D rendering of the native TvL structure (PDB access code 1GYC). *Grey areas* represent residues that were available for mutation in the DLac design while the *green areas* represent amino acids that were left unmutated. *Blue lines* indicate the bonds formed between tightly interacting amino acids (shown as *sticks*). **B** 3-D rendering of the computational design of the DLac protein (The PDB code for DLac can be found in the Supplementary Information). *Green areas* represent amino acids that were left the same as compared to TvL. *Grey areas* indicate amino acids where mutation was allowed, but the wild type amino acid was selected. *Purple areas* indicate amino acids that are different from the TvL structure. *Blue lines* show tightly interacting amino acids. **C** Space filling image of the wild type TvL structure. *Orange amino acids* are polar and *grey amino acids* are non-polar. *Red amino acids* are positively charged while *blue* are negatively charged. **D** Space filling image of the designed structure of the DLac protein, with amino acids color coded the same as in panel **C**. **E** Surface and cavity analysis of the native TvL structure, *blue lines* indicate bonds formed between interacting amino acids on the surface of the protein, (shown as *sticks*). Channels leading to the T1, T2/T3 sites are shown in *green*. Tunnels have been computed using CAVER (Petrek et al. 2006). **F** Surface and cavity analysis of the designed DLac protein, *blue lines* indicate bonds formed between interacting amino acids on the surface of the protein, (shown as *sticks*). Channels leading to the T1, T2/T3 sites are shown in *green*. (Color figure online)

atomic models of the folded and unfolded states are constructed using a molecular mechanics force field, CHARMM22, and a rotamer library. The different computed structures are compared to the final desired structures, and they are scored by their folding free energy which is evaluated as the difference between the free energy in the unfolded and folded states.

The free energy of each state contains contributions arising from electrostatic and van der Waals interactions. The electrostatic interaction is represented by a Coulomb term with dielectric constant $\varepsilon = 8$. The standard Lennard-Jones potential is used to describe van der Waals interactions. An additional term is included that is linear with the solvent accessible area to account for the solvation energy. Energy of a given state is thus computed as the sum of atomic pair-wise contributions (Eq. 1).

$$G = \sum_{i \neq j} \frac{a_{ij}}{r_{ij}^{12}} - \frac{b_{ij}}{r_{ij}^{6}} + \sum_{i \neq j} \frac{q_i q_j}{\varepsilon r_{ij}} + \sum_{i} ASA_i \quad (1)$$

The atomic coefficients are either those of the CHARMM22 force field (a_{ij}, b_{ij}, q_i) or obtained from experimentally measured hydration coefficients for small molecules (σ_i) (Ooi et al. 1987). The unfolded free energy is computed by evaluating the free energy of an ensemble of di-peptides with a composition given by the primary sequence. To rank the sequences, contributions common to all the sequences, like the energetic terms due to the fixed side chains or from the backbone, are neither considered nor computed.

Rotamer library refinement

The side chain conformations in the folded state were initially described using all rotamers with a probability higher than 90% from Dunbrack's backbone dependent library (Dunbrack and Karplus 1993). To enlarge this library, it was further expanded to include rotamers with higher interaction energies. For each pair of rotamers on the surface, a local minimization of their structure in a low-dielectric environment ($\varepsilon = 1$) was performed, so that the search is biased towards H-bond formation (backbone charges are set to zero to further encourage interaction among the rotamers) (Tortosa and Jaramillo 2006). Only pairs with interaction energies lower than -5 kcal/mol were kept. After a new local minimization in the $\varepsilon = 8$ environment, a new refined conformation for each residue of the pair was obtained. To reduce the overwhelming number of rotamers obtained, a final coarse-grained clustering of the conformations was performed and only rotamers differing by more than 25° were stored for further use in the energy computation. A total of 48,150 rotamers were finally obtained.

Combinatorial optimization of protein sequence

A standard Monte Carlo simulated annealing (MCSA) optimization algorithm was utilized and Metropolis was employed as a criterion to accept or reject solutions. The temperature followed an exponential cooling scheme, so that at each of the N iterations decreased from the original value $T_0/R = 1$ kcal/mol to the final value $T_f/R = 0.01$ kcal/mol, according to $T \rightarrow \alpha T_0$. Due to the size of the combinatorial space, a high number of iterations were chosen for the MCSA ($7x10^6$) and 120 independent optimizations were run.

Laboratory materials

The synthetic gene for the designed protein was purchased from DNA2.0 (Menlo Park, CA) as an insert in a standard cloning plasmid (pJ201:10935). Enzymes for DNA cloning and manipulation were from New England Biolabs (Ipswich, MA). *E. coli* strains BL21(DE3) and DH5α as well as the QuikChange Site-directed mutagenesis kit were from Stratagene (La Jolla, CA). Oligonucleotides were obtained from Integrated DNA Technologies (Coralville, IA). The pET-20b(+) expression vector was from Novagen (Darmstadt, Germany). The Gel extraction kit and Miniprep kit were purchased from Qiagen (Valencia, CA). SDS–PAGE gels and dithiothreitol (DTT) were from Invitrogen (Carlsbad, CA). Centricon and Amicon centrifugal filter units were from Millipore (Billerica, MA). Slide-A-Lyzer dialysis cassettes were purchase from Pierce (Rockford, IL). Complete Mini EDTA free protease inhibitor cocktail tablets were obtained from Roche Applied Science (Mannheim, Germany). Isopropyl β-D-1-thiogalactopyranoside (IPTG) from Promega (Madison, Wisconsin). Chromatography columns were from GE Healthcare (Piscataway, NJ). Ampicillin, 2-mercaptoethanol (2-MCE) and phenylmethanesulphonylfluoride (PMSF) from Sigma–Aldrich (St. Louis, MO). All other chemicals were reagent grade or of higher purity and purchased from Sigma–Aldrich (St. Louis, MO) or Fisher Scientific (Fair Lawn, NJ).

Design and cloning of DLac gene

The final computationally designed amino acid sequence (DLac) was modified to include an initiation Met codon. The corresponding DNA sequence was created and optimized for expression in *E. coli* using the Gene Designer software program (Villalobos et al. 2006) and several unique restriction endonuclease sites were included within the sequence, including a C-terminal *Xho*I site (Supplementary Material Fig. 1). The DLac gene was digested from the pJ210: 10935 plasmid using unique *Nde*I and

*Xho*I sites and the fragment was ligated into the pET-20b(+) vector (which contains a C-terminal hexahistidine Tag) using the same restriction sites. This resulted in the addition of 8 amino acids to the C-terminus of the designed DLac sequence (LEHHHHHH). The resulting plasmid (pDLacI) was transformed into *E. coli* strains BL21(DE3) and DH5α and the fidelity of the insertions was verified by DNA sequencing.

Site-directed mutagenesis

Residues 453–455 in the DLac gene that were designed to be involved in T1 and T2/T3 copper binding (analogous to residues 452–454 in TvL) were mutated using site-directed mutagenesis in order to create an inactive negative control protein. The oligonucleotide primer 5′-CGGGCCCGTG GTTCCTGGTCAGCGTCATCGACTTCCACCTGG-3′ and its complementary sequence were used to mutate the amino acid residues 453–455 (HCH) to VSV (encoded by the underlined DNA sequence) in pDLacI. The resulting plasmid pDLac-453HCH/VSV was then transformed into BL21(DE3) cells, and the mutations were verified by DNA sequencing.

Soluble protein expression and purification

Single colonies of *E. coli* harboring either pDLacI or pDLacI-453HCH/VSV were used to inoculate 5 ml LB cultures supplemented with 100 μg/ml ampicillin and after 8 h these were used to inoculate 50 ml cultures which were grown overnight at 37°C to saturation. One liter LB cultures with 100 μg/ml ampicillin were inoculated with 10 ml from the overnight cultures and were grown at 37°C in the presence of 1 mM $CuCl_2$ to an OD_{600} of ~0.9. The temperature was lowered to 25°C and IPTG was added to a final concentration of 1 mM. After 24 h of growth, the cells were pelleted by centrifugation at 8,300×*g* for 10 min and the cell pellets were frozen at −20°C. The thawed pellets were resuspended in 75 ml of ice-cold Resuspension Buffer (20 mM Tris pH 8.0, 500 mM NaCl, 20 mM imidazole, 1 mM DTT, Complete Mini EDTA free protease inhibitor cocktail) and sonicated on ice. The sonicated cells were centrifuged at 25,600×*g* for 30 min. The lysate was loaded onto a 5 ml FF HisTrap immobilized metal affinity chromatography (IMAC) columns using an FPLC apparatus. In a typical run, 150 ml of the lysate (obtained from 2 l of culture) was injected, washed with 5 column volumes (CV) of 90 mM of imidazole (in 20 mM Tris pH 8.0, 500 mM NaCl) and eluted with 5 CV of 500 mM of imidazole in the same buffer. Fractions containing the DLac protein (or mutant), as determined by SDS–PAGE, were pooled and concentrated to 1–2 ml using an Amicon Ultra-15 Centrifugal Filter Unit (30 kDa NMWCO). The concentrated samples were loaded onto a Sephadex 200 PG column for size exclusion chromatography (SEC) equilibrated in an SEC Buffer (50 mM Tris pH 8.0, 500 mM NaCl). Fractions containing DLac (or the mutant) were collected and pooled.

Protein expression and purification from inclusion bodies

One liter cultures with 100 μg/ml ampicillin, inoculated as described above, were grown at 37°C in the presence of 1 mM $CuCl_2$ to an OD_{600} of ~0.6 and IPTG was added to a final concentration of 1 mM. After 4 h of growth at 37°C, the cells were pelleted by centrifugation and frozen as described above. The thawed pellets were sonicated and the lysate was clarified as described above. The supernatant was discarded and the insoluble fractions containing the IBs were resuspended in 30 ml of ice-cold buffer containing 20 mM Tris pH 8.0, 500 mM NaCl, 2 M urea and 2% Triton X-100. The suspension was sonicated on ice. The insoluble material was twice separated by centrifugation at 25,600×*g* for 10 min. The resulting washed IB pellets were resuspended in 50 ml Binding Buffer (20 mM Tris pH 8.0, 500 mM NaCl, 6 M guanidine HCl, 5 mM imidazole and 1 mM 2-MCE) and stirred for 30–60 min at room temperature. The solution was centrifuged at 25,600×*g* for 10 min to remove unsolubilized material.

The solubilized IB proteins were loaded onto IMAC columns for simultaneous purification and refolding of the protein. In a typical run, 100 ml of IB solution was injected to the column and washed with 10 CV of Binding Buffer followed by 10 CV of Washing Buffer (20 mM Tris pH 8.0, 500 mM NaCl, 6 M urea, 20 mM imidazole and 1 mM 2-MCE). Refolding was completed by changing the Washing Buffer to a Refolding Buffer (20 mM Tris pH 8.0, 500 mM NaCl and 20 mM imidazole) using a linear gradient over 30 CV. Experiments were performed with and without the addition of 1 mM $CuCl_2$ to the Refolding Buffer. After washing with 5 CV of Refolding Buffer, the protein was eluted with 5 CV of Elution Buffer (20 mM Tris pH 8.0, 500 mM NaCl and 500 mM imidazole). The fractions with refolded DLac were pooled. The samples were concentrated and further purified by SEC as described above.

Purification of TvL and SLAC laccases

TvL was purchased as a crude protein extract. It was further purified by ion exchange chromatography using a HiPrep16/60 DEAE FF ion exchange chromatography (IEC) column in 10 mM sodium phosphate buffer pH 6.2, and the protein was eluted with a linear NaCl gradient from 0 to 1 M, as previously described (Hudak and Barton

2005). Further purification by SEC was performed as described above in a phosphate buffer (50 mM sodium phosphate, 150 mM NaCl, pH 7.2). The TvL fractions were pooled and concentrated (Amicon Ultra-15 Centrifugal Filter Unit, 30 kDa NMWCO).

The SLAC protein was expressed and purified using IEC and SEC as previously described (Gallaway et al. 2008; Machczynski et al. 2004).

PAGE and mass spectrometry

The purity and proper size of all of the purified protein samples (soluble DLac, refolded DLac, DLac 453HCH/VSV, TvL, SLAC) was verified using SDS–PAGE under reducing conditions. Samples of the soluble DLac and refolded DLac were excised from the gels and mass spectrometry following tryptic digestion was performed at the Columbia University Medical Center Protein Core Facility.

UV–Vis spectroscopy

Absorption scans over a wavelength range from 280 to 800 nm were performed using a SpectraMax M2^e microplate reader (Molecular Devices). Scans of TvL, SLAC, soluble DLac, and refolded DLac samples were compared to blank samples in SEC buffer. All measurements were made in triplicate.

Kinetic activity measurements

A series of experiments were performed to measure the catalytic activity of the DLac proteins following just IMAC purification. A variety of substrates with varying redox potentials were explored over a wide range of pH values. All measurements were performed using a SpectraMax M2^e microplate reader. The buffers, substrates, and other conditions are listed in Supplementary Information Table 1 and Table 2. The enzymes (DLac and DLac 453HCH/VSV) were compared to positive controls (TvL and SLAC) and a no-enzyme negative control. In each case, $\sim$7 μg/ml of enzyme was used, and the absorbance at the appropriate wavelength was monitored at 40°C for several hours. Experiments with some of the substrates were repeated with the addition of 1 mM copper chloride. All measurements were made in at least duplicate.

Atomic absorption spectroscopy

Samples of the soluble DLac and refolded DLac protein in SEC buffer as well as TvL were sent to an outside laboratory (Galbraith Laboratories Inc., Knoxville, TN) for graphite furnace atomic absorption spectroscopy (GFAAS) analysis using a Perkin Elmer AAnalyst 800 GFAA/FLAA Spectrophotometer.

Circular dichroism (CD) spectroscopy

The far-UV CD spectra of SEC purified DLac and TvL samples were measured with a Jasco J-815 CD spectrometer using 0.1 and 0.01 cm optical path length cuvettes. A Peltier-equipped cell holder was used to control the sample temperature at 25°C or 90°C. The CD spectrum of TvL was determined in two buffers (10 mM sodium phosphate buffer at pH 7.2 and in 50 mM Tris buffer containing 500 mM sodium chloride at pH 8.0). The CD spectra of DLac were measured only in the high salt Tris buffer solution due the propensity of the protein to aggregate at lower ionic strengths. Spectrum scans were performed in continuous mode, with a 50 nm/min scan speed, 2 s response time and by accumulating five scans. Mean residue ellipticity values were calculated from ellipticity values using the protein concentrations determined by absorbance at 280 nm and calculated extinction coefficients.

The CD spectrum of TvL in sodium phosphate buffer at pH 8.0 was measured both at 25 and 90°C using the 0.1 cm path length cuvette. The high absorbance of the 50 mM Tris buffer at pH 8.0 containing 500 mM NaCl (which was needed for DLac to stay soluble) necessitated the use of a 0.01 cm path length cuvette due to the high absorbance of the buffer. However, CD measurement at 90°C was not possible in this cuvette due to rapid evaporation of the 25 μl sample. Heat treated DLac and TvL samples were thus prepared by heating sample solutions in microcentrifuge tubes to 90°C for 10 min and cooled to room temperature followed by CD measurement at 25°C.

Chemical denaturation of TvL and DLac was performed by adding urea to a final concentration of 0–8 M in 50 mM Tris buffer, 500 mM NaCl at pH 8.0. The CD spectra of the resulting samples were then measured as described above and then smoothed with a 13 nm window Savitzky–Golay filter using the Spectra Manager (Jasco) software program. The mean residue ellipticity value at 220 nm, the wavelength where denaturation caused the largest change, was used to plot the urea denaturation curves. For the denaturation studies of refolded TvL, the enzyme was prepared by first denaturing the native protein in SEC buffer containing 8 M urea and 1 mM DTT overnight and subsequently dialyzing the sample in a step-wise manner against SEC buffer with decreasing urea concentrations (6 M (for 2 h), 5 M, 4 M, 3 M (for 1 h each), 2 M (for 2 h) and 0 M (overnight)). The resulting refolded TvL was then subjected to denaturation studies using urea as described above. A two-state model was fit to the unfolding data (Greenfield 2006).

Results

Designed DLac sequence

The computational design and optimization process introduced mutations in 253 of the 339 positions that were available for mutation (Fig. 1), and therefore the final DLac primary sequence shares less than 50% sequence homology with the parent TvL sequence (Fig. 2) (PDB file of modeled DLac protein is included in Supplementary Material). Amino acids in the core of the enzyme (within 12 Å radius of the four copper) were left unmutated, and amino acids within 8 Å from the cooper atoms were also forced to have the same structural conformation as observed in the crystal structure. This was done in order to attempt to retain the native copper binding sites and their local environment as well as to retain a tunnel through the protein to the active site of the protein that enables the flux of oxygen and water (Fig. 1).

Comparing the designed and parent TvL sequences, the number of buried positions (considered to be those where the relative accessible solvent area is less than 20%) slightly increased in the DLac sequence from 139 to 146. Pairs of residues were considered to be tightly interacting whenever their combined van der Waals and electrostatic interaction energies were below −5 kcal/mol and they were considered to have a clash whenever their interaction energy was higher than 2.5 kcal/mol. In the published TvL crystal structure, there are 10 tight interactions. Since the library refinement procedure favored rotamers belonging to highly interacting pairs, the DLac sequence introduced 68 of these pairs, which should produce a more extensive H-bond and salt bridge network on the proposed surface of the protein (Fig. 1). The ratio of polar to non-polar amino acids on the surface of the protein changed from 2.4 to 4.5. The wild type protein had 200 amino acids on the surface and among them, 141 were polar, whereas the DLac protein has 164 polar amino acids on its surface. The optimization process also resulted in a lower number of predicted clashes in the proposed protein structure, as the wild type protein has 8 clashes whereas the DLac sequence should only have three. The minimum folding energy of the final DLac protein was estimated to be −2,436.7 kcal/mol compared to −2,115.1 kcal/mol for the wild type TvL, using the same scoring function, indicating that the final folded structure should have increased stability.

```
DLAC  MYIGPQAELEVKEKPLKPAGLELKLIVVKDRLPAPEILGKKGQEFRLRVKVQLRDKKLLKL
1GYC  AIGPAASLVVANAPVSPDGFLRDAIVVNGVFPSPLITGKKGDRFQLNVVDTLTNHTMLKS
       *** *.* * : *:.* *:  . ***:. :*:* * ****:.*:*.*   * ::.:**

DLAC  TSIHWHGFKQKRTNWADGPAFVNQCPIVQGKEFEYKFRVPEQAGTFWYHSHLKTQYCVGL
1GYC  TSIHWHGFFQAGTNWADGPAFVNQCPIASGHSFLYDFHVPDQAGTFWYHSHLSTQYCDGL
      ******** *  ***************..*:.* *.*:**:***********.**** **

DLAC  RGPFLVEDPRDPQRQKWKEWREKALLEIEDLLKQLAQEGPKFPRGLDALLLKGQGRAEEK
1GYC  RGPFVVYDPKDPHASRYDVDNESTVITLTDWYHTAARLGPRFPLGADATLINGLGRSAST
      ****:* **:**: .::.  .*.::: : *  :  *: **:** * ** *::* **: ..

DLAC  PEYRLAKVDVRQGIEKAIALVSISCDPNIKLKVDRLKLEVVEVDGIKVKPLEVKAVEIFA
1GYC  PTAALAVINVQHGKRYRFRLVSISCDPNYTFSIDGHNLTVIEVDGINSQPLLVDSIQIFA
      *   ** ::*::* .  : ********* .:.:*  :* *:*****: :** *.:::***

DLAC  AQRIQAVQWARIEKDILWIKAEPKKGKRDWRRGIAAAYLILEGAPRREPKQREDREERPL
1GYC  AQRYSFVLNANQTVGNYWIRANPNFGTVGFAGGINSAILRYQGAPVAEPTTTQTTSVIPL
      *** . *  *.   .  **:*:*: *. .:  ** :* *  :***  **.  :  .  **

DLAC  EEDRLEPQQDKPVPGKPEPRGVRKWLRLDFEFEGDRFKIQRHEFYPPEEPVLKKIERGKK
1GYC  IETNLHPLARMPVPGSPTPGGVDKALNLAFNFNGTNFFINNASFTPPTVPVLLQILSGAQ
       * .*.*    ****.* * ** * *.* *:*:* .* *:. .* **  *** :*  * :

DLAC  EKKDLDPYGLVDPLPDRADIEIELPAKREAPGKPHPFHLHGHEFAVVKAERKRDLKQERP
1GYC  TAQDLLPAGSVYPLPAHSTIEITLPATALAPGAPHPFHLHGHAFAVVRSAGSTTYNYNDP
        :** * * * *** :: *** ***.  *** ********* ****::  .   : : *

DLAC  VLRDVVSTGEPRQRDQVKIKFRVDNPGPWFLHCHIDFHLDAGFAIVFAAAYEKVWEDVPK
1GYC  IFRDVVSTGTPAAGDNVTIRFQTDNPGPWFLHCHIDFHLEAGFAIVFAEDVADVKAANPV
      ::******* *   *:*.*:*:.****************:********    .*    *

DLAC  PKEWDELCPDEQRKKSEDKLEHHHHHH
1GYC  PKAWSDLCPIYDGLSEANQ
      ** *.:***  :  .. ::
```

Fig. 2 Comparison of DLac and TvL (1GYC) primary sequences. There is less than 50% sequence homology between the two proteins. In order to express and purify the DLac sequence, an N-term methionine was added to the sequence as well as a C-term hexahistidine tag (LEHHHHHH)

Protein expression and purification

The protein was readily overexpressed in *E. coli*, but the majority of the protein was found as insoluble aggregates in IBs. Several parameters were varied in order to improve soluble expression including: temperature, IPTG concentration, copper concentrations, and induction time. The best results were obtained when cultures were moved to 25°C at the time of induction. The expression was not fully optimized, but sufficient quantities of soluble protein were obtained for characterization experiments. Fig. 3 shows the expression levels obtained, and the ratio of protein in the soluble and insoluble fractions. Small quantities of soluble protein were obtained using IMAC and SEC purification.

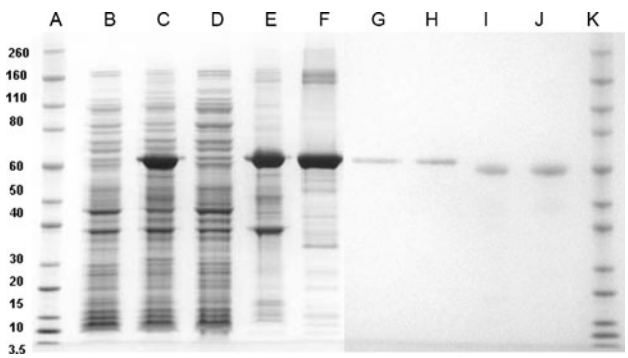

Fig. 3 SDS–PAGE results of the DLac expression and purification. *Lane A* MW Ladder, *Lane B* Whole cell extract of the uninduced cell culture, *Lane C* Whole cell extract of the induced cell culture, *Lane D* Soluble protein fraction from induced cells, *Lane E* Insoluble protein fraction from induced cells solubilized in 8 M urea, *Lane F* Refolded DLac protein from an IMAC column, *Lane G* Soluble DLac protein after SEC, *Lane H* Refolded DLac protein after SEC, *Lane I* Native TvL after SEC column, *Lane J* Denatured and refolded TvL protein after SEC column, *Lane K* MW Ladder

However, the soluble protein precipitated after overnight cold storage unless a large salt concentration (500 mM NaCl) was included in the buffer.

Since the majority of the protein was found in IBs, several refolding protocols were explored. The DLac protein was readily soluble in either 6 M guanidine HCl or 8 M urea, and DTT was added to ensure any disulfides were reduced. The largest amount of protein was obtained when the DLac was refolded on an IMAC column. Copper chloride (1 mM) was added to the Refolding Buffer and/or to the refolded proteins overnight, and following concentration, SEC purification was performed. There were multiple peaks in the SEC chromatograms and a large proportion of the protein was found in the void volume (suggesting aggregation). There was a common peak between the refolded and soluble DLac injections at 73 ml (Fig. 4A, B), and it was this fraction which was used for all subsequent experiments. It was observed that if the time between IMAC refolding and SEC injection was minimized, the largest peak of interest was obtained, which suggests that aggregation increased over time.

Native TvL did not share the same retention time in the SEC chromatograms as either the soluble or refolded DLac protein. However, the denatured and refolded TvL did share the same retention time as the DLac proteins (Fig. 4C, D).

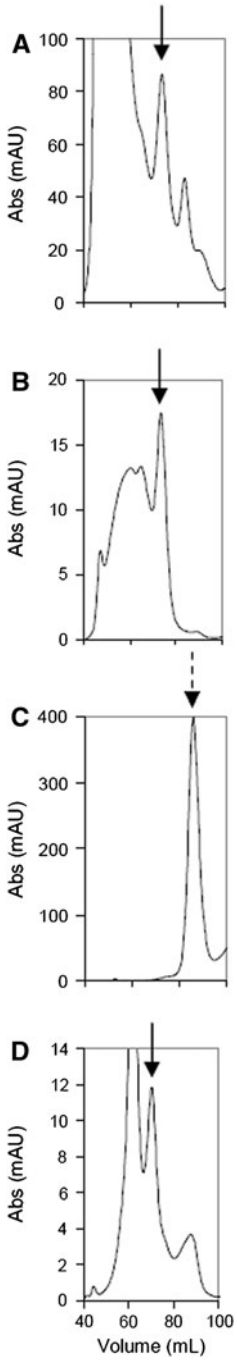

Fig. 4 SEC chromatograms for DLac and TvL. Panel **A** Soluble DLac protein, Panel **B** Refolded DLac protein, Panel **C** Native TvL protein, Panel **D** Denatured and refolded TvL (with the refolding occurring in the SEC column). For the DLac and refolded TvL proteins, the peak that eluted at ~73 ml was used (*black arrows*). For the native TvL, the active protein eluted later (*dashed arrow*)

Mass and copper content

Trypsin digested samples from an SDS–PAGE gel band showed that both the soluble and refolded DLac protein peaks give rise to the same digestion pattern, which is consistent with the theoretical digestion pattern of the computationally designed protein (Supplementary Material Table 3).

The blue T1 copper, if present in the protein, is characterized by an absorption peak at ~600 nm. For the TvL and SLAC proteins, a wavelength scan reveals the correct distinctive peak. Neither the solubly expressed DLac protein nor the refolded DLac protein showed a peak in this region (Fig. 5). This was independent of whether the

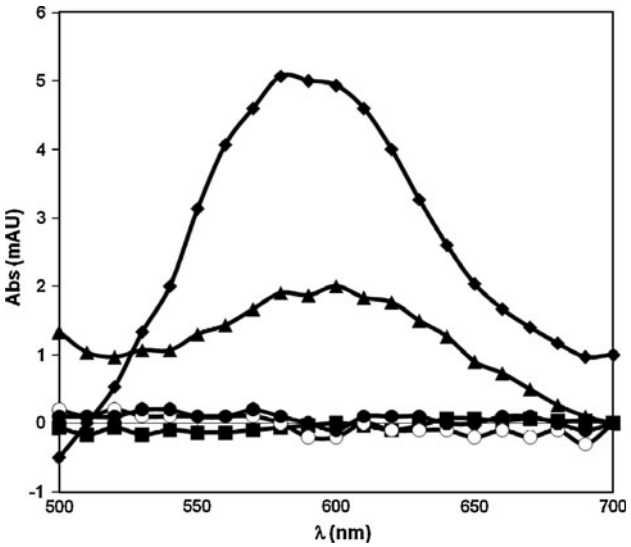

Fig. 5 Spectroscopic scan for the copper T1 site of the soluble DLac (*shaded square*), refolded DLac in the absence of copper (*shaded circle*), refolded DLac in presence of copper (*open circle*), TvL (*shaded triangle*), SLAC (*shaded diamond*). All samples were measured after SEC purification

proteins were incubated with 1 mM copper chloride before subjecting them to SEC.

GFAAS analysis was unable to detect copper ions in either the soluble or refolded DLac proteins. Copper was present in the expected amount of 4 copper ions per molecule for the TvL sample.

Kinetic activity measurements

Although the DLac protein did not appear to bind a significant amount of copper, kinetic assays were still performed in case small amounts of the enzyme had folded properly. In order to maximize the chances of measuring kinetic activity, proteins purified by IMAC (and not SEC) were used. The IMAC samples would include the protein correctly identified by mass spectrometry, as well as any other forms of the protein that might behave differently (and thus separated out) by the SEC column step. Since some of the substrates can autooxidize, the DLac 453HCH/VSV mutant, purified identically, was used as a negative control.

The TvL protein was active at a pH range from 4 to 8, oxidizing all substrates in Supplementary Material Table 2 as observed in colorimetric assays. SLAC was able to oxidize DMP, guaiacol, MBA and SYG at pH values ranging from 7 to 10, with higher activity usually at pH 8–9. SLAC also was able to oxidize ABTS at a pH range of 4–7, similar to that of TvL. SLAC did not appear to be able to oxidize DMBA.

At pH values ranging of 2–11 for all of the substrates in Supplementary Material Table 2, the DLac protein (either solubly expressed or refolded) and the triple mutant DLac 453HCH/VSV exhibited identical colorimetric results, indicating a lack of catalytic activity under these experimental conditions.

DLac and TvL conformation

CD spectroscopy was used to study the protein conformation of DLac while comparing it to that of TvL (Fig. 6). The CD spectra of both the refolded DLac and the DLac from the soluble fraction had the same shape with a minimum at about 208 nm (Fig. 6A, B). The spectra were deconvoluted and secondary structure compositions of 9% alpha helix, 33% beta sheet, 19% turns, and 35% unordered were calculated for the solubly expressed form of DLac, and 9, 35, 18 and 39% respectively for the refolded DLac. The CD spectrum of TvL had a maximum at 196 nm and a minimum at around 216 nm resembling a structure that is rich in beta sheets (Fig. 6C). Estimation of secondary structure contents by deconvolution yields 5% helix, 41% beta sheet and 50% other structures, which are in agreement with the TvL X-ray crystal structure(Piontek et al. 2002) that has 11% helical, 37% beta sheet, and 52% other structures according to the secondary structure assignments from the program DSSP (Kabsch and Sander 1983).

When the DLac samples were dissolved in 8 M urea, a significant change in the CD specta was observed as there

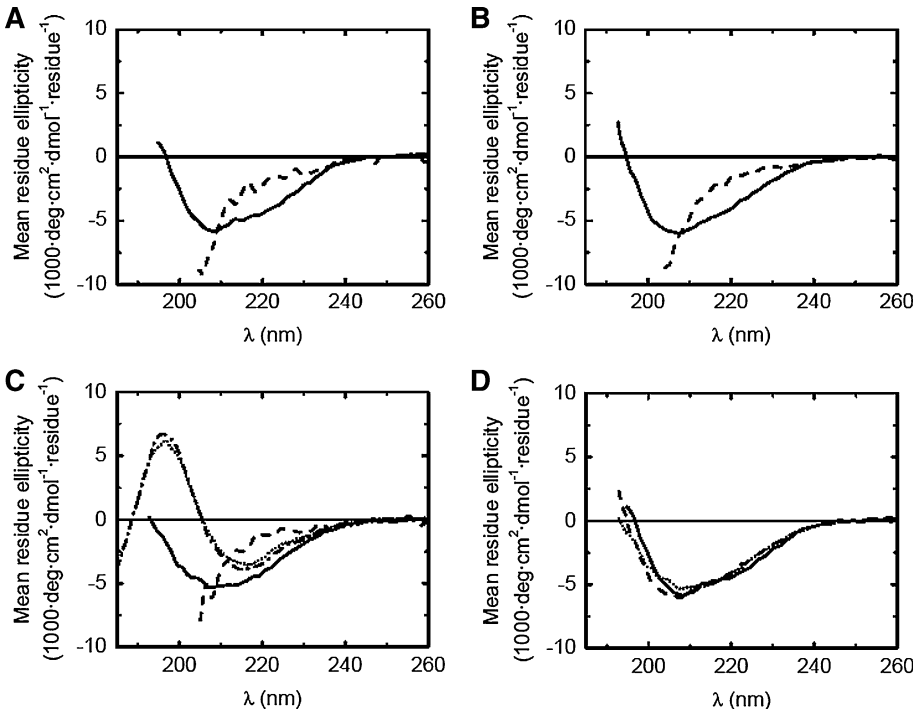

Fig. 6 CD spectra of SEC purified DLac and TvL laccases. **A** Soluble DLac in 0 M (*solid line*) and 8 M (*dashed with dotted line*) urea. **B** Refolded DLac labeled the same as Panel **A**. **C** TvL in 0 M (*dotted line*) and 8 M (*dashed with dotted line*) urea, and TvL that was heat denatured and cooled (*solid line*). The samples were measured in a high ionic strength buffer (as described in "Materials and methods") and for comparison the CD spectrum of TvL in 10 mM Na-phosphate buffer at pH 7.2 (*dashed with dotted line*) is shown in panel **C**. **D** Direct comparison of soluble (*solid line*) and refolded (*dashed with dotted line*) DLac spectra with that of heat denatured TvL (*dotted line*)

was an increase in unordered structure from 35 to 47% for the solubly expressed DLac and an increase from 39 to 52% for the refolded DLac. When the TvL was heat-treated at 90°C and cooled to 25°C, a CD spectrum (10% helical, 34% beta sheet) was observed that was very similar to what was seen with the DLac samples. After the heat treatment, TvL did not exhibit detectable enzymatic activity when tested with ABTS. In order to further compare TvL with the DLac samples, the TvL was chemically denatured and refolded using step-vise dialysis to mimic the DLac refolding protocol. Again, the CD spectrum of the denatured and refolded TvL was very similar to the refolded DLac protein as shown in Fig. 6C and D.

The thermodynamic stability of DLac was investigated at high ionic strength to prevent the aggregation of the protein. The refolded DLac, denatured and refolded TvL, and native TvL were denatured in urea and the unfolding was followed using CD spectroscopy. All three proteins had similar weakly sigmoidal unfolding curves (Fig. 7) suggesting similar stabilities. It is possible that the high (500 mM NaCl) ionic strength destabilizes the proteins. The free energy of folding for DLac and refolded TvL were −1.8 kcal/mol and −2.2 kcal/mol, respectively. The unfolding–folding transition of native TvL is not reversible under the conditions used, and thus the thermodynamic analysis cannot be applied.

Discussion

We have shown how computational design could be used to engineer a large protein with little sequence similarity to any known sequence. Although current computational methods are still far from getting a folded and active protein of this size, our attempt highlights the important factors to consider in the design process. There are many examples of laccase and laccase-like proteins identified from prokaryotic organisms, and several of these have been recombinantly overexpressed in *E. coli*. Since the fungal laccase proteins generally have 3-D structures and active site architectures that are similar to the prokaryotic laccase enzymes, there is no obvious reason why the structure of a fungal laccase cannot similarly be expressed in a bacterial host. In fact, it has recently been reported that a fungal laccase (from *C. bulleri*) has been heterologously expressed in *E. coli* in a catalytically active form (Salony et al. 2008).

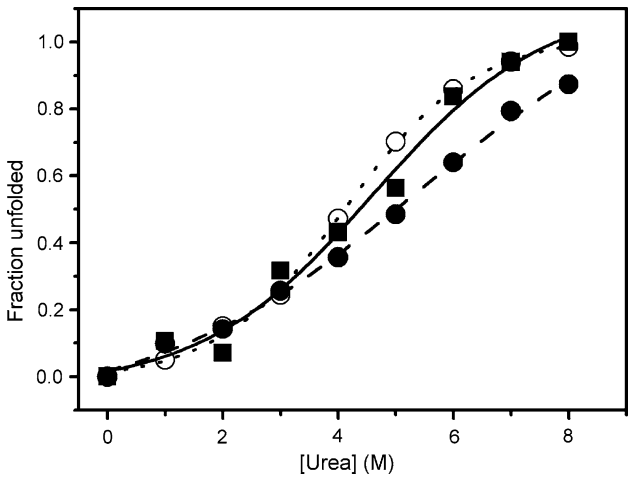

Fig. 7 Urea denaturation of refolded DLac (*shaded square*), denatured and refolded TvL (*open circle*) and native TvL (*shaded circle*) proteins as measured by the change in CD signal at 220 nm. Refolded TvL and refolded DLac showed reversible folding–unfolding and thus a two-state unfolding model was fit to the refolded DLac (*solid line*) and refolded TvL (*dotted line*) data (see "Materials and methods"). The unfolding of native TvL (*dashed line*) is not reversible

Computational protein design has been successfully used to identify primary amino acid sequences that will adopt a predetermined tertiary protein fold (Jaramillo et al. 2002; Ogata et al. 2003; Wernisch et al. 2000). We set out to use this approach to determine a primary sequence that would fold into the tertiary structure of the TvL glycoprotein. Ideally this novel designed protein would exhibit increased stability, and it would retain the high redox potential and activity of the parental protein which would be beneficial for a variety of applications.

Laccase enzymes follow a bi-bi ping-pong mechanism such that the substrate is oxidized by the T1 copper atom, the electron is conducted through the core of the protein to the T2/T3 site, and four electrons are used to reduce dioxygen to water. It is clear that the proper orientation and local environment around the copper atoms is essential for catalysis and therefore the computational design strategy left all amino acids within 12 Å of the copper atoms unchanged. The final design allowed 339 of the 499 amino acids in the protein to be mutated making this one of the largest reported computational protein design efforts reported. The final designed DLac sequence had less than 50% sequence identity with that of parent TvL (Fig. 2), but it was correctly recognized by sequence alignment servers like BLAST to be a copper-oxidase.

One of the most interesting features of the new sequence is the substantial increase in tightly interacting amino acid pairs that were designed to form on the surface of the properly folded protein (Fig. 1C, D). This would result in a more extensive H-bond and salt bridge network on the surface of the properly folded protein, ideally leading to increased stability of the protein. This is a result of the computational methodology which tends to favor electrostatic interactions at the surface and not at the core (Jaramillo et al. 2002), but this increased polar network has not been found to be incompatible with folding (Suarez and Jaramillo 2009). The calculated theoretical pI of the DLac protein using ProtParam (Walker 2005) was 9.26 as

compared to 5.87 for TvL (and an experimental value of 3.5 has been reported (Piontek et al. 2002)). The large change in the charges of the protein sequence may have contributed to the observed propensity for the polypeptide to aggregate, as it was found that in order for the expressed DLac protein to remain soluble, a high ionic strength buffer was required (500 mM NaCl). This is contrast to what was observed for the denatured and refolded native TvL protein, which did not require high salt to remain soluble.

Transformation of the designed gene in *E. coli* resulted in the high expression of protein that was preferentially located in IBs and several established methods were used to attempt to refold the insoluble protein. Protocols were developed to refold the DLac protein as well as to purify the small amount of solubly expressed DLac protein. Once purified, these samples behaved identically in every test performed, suggesting that they are in a similar conformation. The absence of the blue T1 copper combined with the lack of copper detected using GFAAS is consistent with lack of apparent kinetic activity of the enzymes under the experimental conditions tested.

In order to further explore whether the DLac protein was properly folded, both SEC and CD measurements were performed. The SEC experiments demonstrated that both DLac samples as well as the denatured and refolded TvL elute at the same retention time, while the native (active) TvL elutes at a later time (Fig. 4). These results are consistent with the CD experiments where it was shown that both DLac samples have similar secondary structure contents. These spectra are very different from the native TvL spectrum, but are very similar to the spectrum of the chemically denatured and refolded TvL (which is no longer active) (Fig. 6). A similar CD spectrum has been also reported for another fungal laccase that was heat-treated (López-Cruz et al. 2006). Taken together these results suggest that the designed DLac polypeptide is able to adopt a similar soluble structure in both the cellular milieu, and following denaturation and refolding of protein accumulated in IBs. This appears to be a molten globule conformation that is very similar to the denatured and refolded TvL.

The weakly sigmoidal unfolding curves (Fig. 6) indicate that the resistance towards chemical denaturation is similar in all 3 protein samples. Moreover, it suggests that the proteins studied have a low cooperativity in unfolding and thus a non two-state folding-unfolding process is expected. TvL is a comparatively large protein with three domains and unfolding may involve significant accumulation of intermediate states. Furthermore, the high ionic strength may destabilize the structure of native TvL. The folding–unfolding process for TvL is not reversible and thus thermodynamic analysis cannot be applied. However, the folding and unfolding of DLac and the in vitro folded TvL are reversible. A two-state model, that assumes no accumulation of an intermediate state, fit well to the unfolding curves. An unfolding free energy change of 1.8 kcal/mol for DLac and 2.2 kcal/mol for in vitro folded TvL was obtained. Thus, the thermodynamic stabilities of the two proteins are comparable and relatively low.

We speculate that this refolded structure could be a molten globular stable folding intermediate that cannot proceed to the native structure. This hypothesis is based on the observations that the protein can adopt a soluble 3-D structure; the structure is similar to the denatured and refolded TvL but not the native TvL, the protein does not bind copper and is inactive, and the protein can be reversibly unfolded with a low thermodynamic stability. One explanation for this could be limitations in copper incorporation into the DLac protein. In addition to their catalytic significance, bound coppers have been reported to have an important structural role in multicopper oxidases (Durão et al. 2006, 2008; Sedlak and Wittung-Stafshede 2007). It seems that native glycosylations are not sufficient to enable correct refolding as the native TvL sample (that is fully glycosylated) could not be refolded from the molten globular structure.

The successful chemical denaturation and refolding of some other laccases has previously been reported in the literature. The *B. subtilis* endospore coat protein, CotA, which is a prokaryotic laccase-like protein when expressed in *E. coli* was initially found to accumulate in IBs and initial refolding experiments were similarly unsuccessful (Martins et al. 2002), but conditions were obtained to enable the production of a soluble and functional CotA protein.(Sakasegawa et al. 2006) And, the fungal laccase from *C. bulleri* was successfully refolded and found to be active after subjecting it to several denaturating and metal chelating agents.(Salony et al. 2008) Unlike TvL, this laccase does not contain cysteine residues which may have simplified the refolding process for this protein.

Computational protein design provides an opportunity to test our understanding of protein folding rules and to validate these assumptions experimentally. Since proteins with structures very similar to TvL can fold into a functional form in prokaryotic hosts, there is no fundamental reason why a protein with an identical structure to TvL cannot similarly be expressed and folded in this environment. Therefore, future experiments will focus on determining if sequence features found in the prokaryotic laccases are missing in the DLac design, and this will help to better elucidate the folding pathway utilized by TvL and related multi-copper oxidases. In addition, future experiments will also explore the expression of the DLac protein in a eukarotic host, such as yeast, to see if this can enable functional folding and expression.

To the best of our knowledge, this is the largest protein that has ever been designed using these computational

techniques. Laccases are multi-domain proteins generally comprised of 3 cupredoxin-like domains each with a greek key β-barrel topology (Giardina et al. 2010; Skalova et al. 2009). These domains likely evolved to be stable independent of the entire laccase protein. This suggests that future design efforts may also benefit from the independent modular design of the domains in conjunction with the design of the protein as a whole.

Conclusion

We have successfully expressed a computational designed laccase based on a fungal protein backbone in a bacterial host. The DLac protein is one of the largest computationally designed proteins expressed in an organism to date. The amount of soluble protein compared to the insoluble aggregates was small, but enough to conduct preliminary characterization experiments.

The soluble and refolded DLac proteins were apparently identical and neither of the proteins was found to be structurally comparable to the native TvL protein. Instead, the DLac proteins were apparently structurally similar to denatured and refolded TvL, which suggests that the native TvL requires a folding pathway that cannot be replicated following chemical denaturation. Future attempts to design this protein will likely require the incorporation of a folding pathway, which may be included by considering the stabilization of intermediate structures, although there are no templates for these presently available.

Copper ions were not detected in the DLac protein samples and the lack of copper is consistent with the observed lack of kinetic activity. These results suggest that in the absence of the fungal cellular context, both DLac and TvL fold into the same soluble but inactive structure. These structures have a similar stability according to the unfolding curves, thus suggesting that the proteins are in a molten globule-like state.

Acknowledgments The authors would like to acknowledge the financial support of a Joint Research Project Award from the Alliance Program involving Columbia University and École Polytechnique awarded to S. B and A. J. Financial support was also provided by an AFOSR MURI award (FA9550-06-1-0264) to S. B. D. J. G. acknowledges support from Merck & Co., Inc. and G. S. Z. from the Academy of Finland and the Alfred Kordelin Foundation. A. J. acknowledges support from FP6-NEST-043340 (BioModularH2), FP7-ICT-043338 (Bactocom), FP7-KBBE-212894 (Tarpol), the ATIGE-Genopole and the Fondation pour la Recherche Medicale. A. J. also acknowledges the HPC-Europa program (RII3-CT-2003-506079) and the BSC for supercomputing time. The authors also thank Dr. Ian Wheeldon for the expression and purification of the SLAC protein.

References

Alexandre G, Zhulin IB (2000) Laccases are widespread in bacteria. Trends Biotechnol 18:41–42

Barton SC, Gallaway J, Atanassov P (2004) Enzymatic biofuel cells for implantable and microscale devices. Chem Rev 104: 4867–4886

Brissos V, Pereira L, Munteanu FD, Cavaco-Paulo A, Martins LO (2009) Expression system of CotA-laccase for directed evolution and high-throughput screenings for the oxidation of high-redox potential dyes. Biotechnol J 4:558–563

Bulter T, Alcalde M, Sieber V, Meinhold P, Schlachtbauer C, Arnold FH (2003) Functional expression of a fungal laccase in *Saccharomyces cerevisiae* by directed evolution. Appl Environ Microbiol 69:987–995

Claus H (2003) Laccases and their occurrence in prokaryotes. Arch Microbiol 179:145–150

Dunbrack RL, Karplus M (1993) Backbone-dependent rotamer library for proteins—application to side-chain prediction. J Mol Biol 230:543–574

Durão P, Bento I, Fernandes A, Melo E, Lindley PF, Martins L (2006) Perturbations of the T1 copper site in the CotA laccase from *Bacillus subtilis*: structural, biochemical, enzymatic and stability studies. J Biol Inorg Chem 11:514–526

Durão P, Chen Z, Fernandes A, Hildebrandt P, Murgida D, Todorovic S, Pereira M, Melo E, Martins L (2008) Copper incorporation into recombinant CotA laccase from *Bacillus subtilis*: characterization of fully copper loaded enzymes. J Biol Inorg Chem 13:183–193

Endo K, Hayashi Y, Hibi T, Hosono K, Beppu T, Ueda K (2003) Enzymological characterization of EpoA, a laccase-like phenol oxidase produced by *Streptomyces griseus*. J Biochem (Tokyo) 133:671–677

Festa G, Autore F, Fraternali F, Giardina P, Sannia G (2008) Development of new laccases by directed evolution: functional and computational analyses. Proteins 72:25–34

Gallaway J, Wheeldon I, Rincon R, Atanassov P, Banta S, Barton SC (2008) Oxygen-reducing enzyme cathodes produced from SLAC, a small laccase from Streptomyces coelicolor. Biosens Bioelectron 23:1229–1235

Gelo-Pujic M, Kim HH, Butlin NG, Palmore GT (1999) Electrochemical studies of a truncated laccase produced in *Pichia pastoris*. Appl Environ Microbiol 65:5515–5521

Giardina P, Faraco V, Pezzella C, Piscitelli A, Vanhulle S, Sannia G (2010) Laccases: a never-ending story. Cell Mol Life Sci 67:369–385

Glykys DJ, Banta S (2009) Metabolic control analysis of an enzymatic biofuel cell. Biotechnol Bioeng 102:1624–1635

Greenfield NJ (2006) Determination of the folding of proteins as a function of denaturants, osmolytes or ligands using circular dichroism. Nat Protoc 1:2733–2741

Gupta N, Farinas ET (2009) Narrowing laccase substrate specificity using active site saturation mutagenesis. Comb Chem High Throughput Screen 12:269–274

Hudak NS, Barton SC (2005) Mediated biocatalytic cathode for direct methanol membrane-electrode assemblies. J Electrochem Soc 152:A876–A881

Jaramillo A, Wernisch L, Hery S, Wodak SJ (2002) Folding free energy function selects native-like protein sequences in the core but not on the surface. Proc Natl Acad Sci USA 99:13554–13559

Kabsch W, Sander C (1983) Dictionary of protein secondary structure—pattern-recognition of hydrogen-bonded and geometrial features. Biopolymers 22:2577–2637

Kunamneni A, Camarero S, Garcia-Burgos C, Plou FJ, Ballesteros A, Alcalde M (2008) Engineering and Applications of fungal laccases for organic synthesis. Microb Cell Fact 7:32

Li X, Wei Z, Zhang M, Peng X, Yu G, Teng M, Gong W (2007) Crystal structures of *E. coli* laccase CueO at different copper concentrations. Biochem Biophys Res Commun 354:21–26

López-Cruz JI, Viniegra-Gonzalez G, Hernández-Arana A (2006) Thermostability of native and pegylated *Myceliophthora thermophila* laccase in aqueous and mixed solvents. Bioconjug Chem 17:1093–1098

Machczynski MC, Vijgenboom E, Samyn B, Canters GW (2004) Characterization of SLAC: a small laccase from Streptomyces coelicolor with unprecedented activity. Protein Sci 13:2388–2397

Madzak C, Mimmi MC, Caminade E, Brault A, Baumberger S, Briozzo P, Mougin C, Jolivalt C (2006) Shifting the optimal pH of activity for a laccase from the fungus Trametes versicolor by structure-based mutagenesis. Protein Eng Des Sel 19:77–84

Martins LO, Soares CM, Pereira MM, Teixeira M, Costa T, Jones GH, Henriques AO (2002) Molecular and biochemical characterization of a highly stable bacterial laccase that occurs as a structural component of the *Bacillus subtilis* endospore coat. J Biol Chem 277:18849–18859

Mayer AM, Staples RC (2002) Laccase: new functions for an old enzyme. Phytochemistry 60:551–565

Nakamura K, Go N (2005) Function and molecular evolution of multicopper blue proteins. Cell Mol Life Sci 62:2050–2066

Ogata K, Jaramillo A, Cohen W, Briand JP, Connan F, Choppin J, Muller S, Wodak SJ (2003) Automatic sequence design of major histocompatibility complex class I binding peptides impairing CD8(+) T cell recognition. J Biol Chem 278:1281–1290

Ooi T, Oobatake M, Nemethy G, Scheraga HA (1987) Accesible surface-areas as a measure of the thermodynamic parameters of hydration of peptides. Proc Natl Acad Sci USA 84:3086–3090

Petrek M, Otyepka M, Banas P, Kosinova P, Koca J, Damborsky J (2006) CAVER: a new tool to explore routes from protein clefts, pockets and cavities. BMC Bioinform 7:316

Piontek K, Antorini M, Choinowski T (2002) Crystal structure of a laccase from the fungus Trametes versicolor at 1.90-A resolution containing a full complement of coppers. J Biol Chem 277:37663–37669

Riva S (2006) Laccases: blue enzymes for green chemistry. Trends Biotechnol 24:219–226

Rodgers CJ, Blanford CF, Giddens SR, Skamnioti P, Armstrong FA, Gurr SJ (2010) Designer laccases: a vogue for high-potential fungal enzymes? Trends Biotechnol 28:63–72

Sakasegawa S, Ishikawa H, Imamura S, Sakuraba H, Goda S, Ohshima T (2006) Bilirubin oxidase activity of Bacillus subtilis CotA. Appl Environ Microbiol 72:972–975

Sakurai T, Kataoka K (2007) Basic and applied features of multicopper oxidases, CueO, bilirubin oxidase, and laccase. Chem Rec 7:220–229

Salony GargN, Baranwal R, Chhabra M, Mishra S, Chaudhuri TK, Bisaria VS (2008) Laccase of *Cyathus bulleri*: structural, catalytic characterization and expression in *Escherichia coli*. Biochim Biophys Acta 1784:259–268

Sedlak E, Wittung-Stafshede P (2007) Discrete roles of copper ions in chemical unfolding of human ceruloplasmin. Biochemistry 46:9638–9644

Shleev S, Tkac J, Christenson A, Ruzgas T, Yaropolov AI, Whittaker JW, Gorton L (2005) Direct electron transfer between copper-containing proteins and electrodes. Biosens Bioelectron 20:2517–2554

Skalova T, Dohnalek J, Ostergaard LH, Osteryaard PR, Kolenko P, Duskova J, Stepankova A, Hasek J (2009) The structure of the small laccase from *Streptomyces coelicolor* reveals a link between laccases and nitrite reductases. J Mol Biol 385:1165–1178

Solomon EI, Sundaram UM, Machonkin TE (1996) Multicopper oxidases and oxygenases. Chem Rev 96:2563–2606

Suarez M, Jaramillo A (2009) Challenges in the computational design of proteins. J R Soc Interface 6(Suppl 4):S477–S491

Tortosa P, Jaramillo A (2006) Active sites by computational protein design. In: Proceedings of the II BIFI 2006 international conference, pp 96–101

Villalobos A, Ness JE, Gustafsson C, Minshull J, Govindarajan S (2006) Gene designer: a synthetic biology tool for constructing artificial DNA segments. BMC Bioinform 7:285

Walker JM (2005) The proteomics protocols handbook. Humana Press, Totowa

Wernisch L, Hery S, Wodak SJ (2000) Automatic protein design with all atom force-fields by exact and heuristic optimization. J Mol Biol 301:713–736

Wheeldon IR, Gallaway JW, Barton SC, Banta S (2008) Bioelectrocatalytic hydrogels from electron-conducting metallopolypeptides coassembled with bifunctional enzymatic building blocks. Proc Natl Acad Sci U S A 105:15275–15280

Xu F (1997) Effects of redox potential and hydroxide inhibition on the pH activity profile of fungal laccases. J Biol Chem 272:924–928

Xu F, Shin W, Brown SH, Wahleithner JA, Sundaram UM, Solomon EI (1996) A study of a series of recombinant fungal laccases and bilirubin oxidase that exhibit significant differences in redox potential, substrate specificity, and stability. Biochim Biophys Acta 1292:303–311

Xu F, Berka RM, Wahleithner JA, Nelson BA, Shuster JR, Brown SH, Palmer AE, Solomon EI (1998) Site-directed mutations in fungal laccase: effect on redox potential, activity and pH profile. Biochem J 334(Pt 1):63–70

Xu F, Palmer AE, Yaver DS, Berka RM, Gambetta GA, Brown SH, Solomon EI (1999) Targeted mutations in a Trametes villosa laccase. Axial perturbations of the T1 copper. J Biol Chem 274:12372–12375