Published in final edited form as:

Chemosphere. 2010 November ; 81(7): 897–903. doi:10.1016/j.chemosphere.2010.08.001.

QSAR-Assisted Design of an Environmental Catalyst for Enhanced Estrogen Remediation

Lisa M. Colosi^{1,*}, Qingguo Huang², and Walter J. Weber Jr³

¹Department of Civil & Environmental Engineering, University of Virginia, Charlottesville, VA

²Department of Crop & Soil Sciences, University of Georgia, Griffin, GA

³Department of Chemical Engineering, University of Michigan, Ann Arbor, MI

Abstract

A quantitative structure-activity relationship (QSAR) was used to streamline redesign of a model environmental catalyst, horseradish peroxidase (HRP), for enhanced reactivity towards a target pollutant, steroid hormone 17 β -estradiol. This QSAR, embodying relationship between reaction rate and intermolecular binding distance, was used *in silico* to screen for mutations improving enzyme reactivity. Eight mutations mediating significant reductions in binding distances were expressed in *Saccharomyces cerevisiae*, and resulting recombinant HRP strains were analyzed to determine Michaelis-Menten parameters during reaction with the target substrate. Enzyme turnover rate, $\ln(k_{CAT})$, exhibited inverse relationship with model-predicted binding distances ($R^2 = 0.81$), consistent with the QSAR. Additional analysis of native substrate degradation by selected mutants yielded unexpected increases in $\ln(k_{CAT})$ that were also inversely correlated ($R^2 = 1.00$) with model-predicted binding distances. This suggests that the mechanism of improvement comprises a nonspecific “opening up” of the active site such that it better accommodates environmental estrogens of any size. The novel QSAR-assisted approach described herein offers specific advantages compared to conventional design strategies, most notably targeting an entire class of pollutants at one time and a flexible hybridization of benefits associated with rational design and directed evolution. Thus, this approach is a promising tool for improving enzyme-mediated environmental remediation.

Keywords

QSAR-assisted design; protein engineering; 17 β -estradiol; horseradish peroxidase

Introduction

The pronounced occurrence of “feminization” and other endocrine abnormalities at sites downstream from municipal wastewater treatment plants (WWTPs) was first noted in 1994 (Purdum et al., 1994). Since then, it’s been demonstrated that these effects are caused by a diverse array of “environmental estrogens,” which, by virtue of their structural similarity to steroid hormones, can elicit a response from cellular estrogen receptors. A key molecular

© 2010 Elsevier Ltd. All rights reserved

*Corresponding Author – lcolosi@virginia.edu, Tel: (434) 924-7961, Fax: (434) 982-2951; Address: B228 Thornton Hall (Civil Engineering); PO Box 400742 Charlottesville, VA 22904-4742.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

structure responsible for these effects is the hydroxylated aromatic ring (Fang et al., 2001), such that virtually all phenolic chemicals exhibit some degree of estrogenic toxicity (Nishihara et al., 2000). These chemicals are ubiquitous in the global environment, persistent, bioaccumulative, and capable of eliciting physiological responses at exceedingly low concentrations (0.1~10 ng/L) (Johnson and Sumpter, 2001).

In light of their significant potential to cause adverse health effects and their resistance to conventional drinking water and wastewater treatments (Johnson and Sumpter, 2001), environmental estrogens have become a critical treatment and research priority. This had led to interest in various oxidative treatments (e.g. photolysis, sonolysis, Fenton chemistry, etc.), with some particular emphasis on the promise of enzyme-mediated pollution abatement. Compared to conventional wastewater treatment and other oxidative treatments, the use of extracellular enzymes for removal of estrogens from water has several particularly appealing benefits (Gianfreda and Rao, 2004), including: 1) the unique specificity of enzyme reactions, which enables rapid reaction even at exceedingly low substrate concentrations; 2) the diverse array of naturally-occurring enzyme catalysts exhibiting native reactivity towards various xenobiotic substrates; and 3) the potential usefulness of protein design to engineer specially-tailored enzymes for individual contaminants or classes of particular regulatory concern. For the particular case of environmental estrogens, “peroxidase” enzymes (including horseradish peroxidase, manganese peroxidase, haloperoxidase and others) possess both demonstrated detoxification abilities and real potential for effective protein engineering (Auriol et al., 2007; Caza et al., 1999; Colosi et al., 2007a, 2007b; Colosi et al., 2009).

Directed evolution and rational design are two general classes of protein design strategies that have been utilized most widely to date (Ang et al., 2005). Directed evolution refers to an iterative process of random mutagenesis, high-throughput screening, and genetic recombination to identify and produce enzyme mutants with enhanced performance compared to the wild type (Arnold and Volkov, 1999). In contrast, rational design strategies rely on mechanistic computational models to understand interactions between the enzyme and its substrate. They then make use of site-directed mutagenesis to target specific changes in enzyme structure such that the performance of the resulting mutant is improved compared to the wild type (Ang et al., 2005).

With respect to designing enzymes that better degrade pollutants, both types of protein engineering strategies have been used with some success. Several highlights include rational design experiments by Carmichael and Wong (2001) and Harford-Cross et al. (2000) in which bacterial P450 enzymes were mutated at their active sites to enhance oxidation rates of several important polycyclic aromatic hydrocarbons (PAHs). The extent of degradation rate increases ranged from 2-fold to 200-fold. Similarly, directed evolution studies by Joo et al. (1999) and Salazar et al. (2003) resulted in up to 20-fold increases in naphthalene oxidation rate by P450 and increased thermostability of this enzyme, respectively. Investigations by DiSioudi et al. (1999) and Watkins et al. (1997) utilized rational design to create mutant organophosphorus hydrolases (OPHs) with increased catalytic turnover rate towards non-native substrates, including nerve agent VX and the neurotoxic insecticide methyl parathion. Additionally, two directed evolution investigations by Cho et al. (2002; 2006) yielded up to 25-fold increases in degradation rate of organophosphorus neurotoxins such as parathion and coumaphos.

The estrogens’ potency and ubiquity in the water supply make them a more immediate, if less acutely toxic, threat to human health than PAHs and neurotoxins targeted in previous protein design studies. This makes protein engineering to improve enzyme-mediated estrogen removal, the goal of this research, an important objective. In contrast to the protein engineering investigations referenced above, this work leverages a previously existing quantitative structure-activity relationship (QSAR) for HRP reactions with estrogenic phenols. Thus, we

refer to our protein design strategy as QSAR-assisted catalyst design. We expect that QSAR-assisted design strategies offer a compelling mix of some benefits from both rational design and directed evolution.

It was previously reported that HRP turnover rate, $\ln(k_{\text{CAT}})$, could be accurately predicted as a function of two independent predictors: 1) the energy of a phenolic substrate's highest occupied molecular orbital (E_{HOMO}); and, 2) the characteristic binding distance between that substrate's phenolic proton and δN on HIS42's imidazole ring (Colosi et al., 2006). The excellent degree of correlation exhibited by this QSAR model was indirect evidence for the following hypothesis: An estrogenic phenol's HRP removal rate is intrinsically determined by its E_{HOMO} value unless the substrate is too sterically hindered to facilitate optimal binding; in which case it will manifest retardation relative to its maximum potential. This hypothesis was subsequently validated *in silico* using literature measurements of 2-methoxyphenol degradation rates for several HRP mutants in conjunction with probabilistic binding simulations for each selected mutant. It was thus demonstrated that reduction in average binding distance between HRP and an estrogenic phenol mediates a linearly proportional increase in $\ln(k_{\text{CAT}})$ (Colosi et al., 2007b).

The current report summarizes successful redesign of a model oxidative enzyme, horseradish peroxidase (HRP), for enhanced reactivity towards the endogenous estrogen, 17β -estradiol, and its structural analogs. Although this is important in and of itself, this paper also points to the usefulness of what we call "QSAR-assisted" protein design"; i.e., the use of computational simulation, as guided by an empirical quantitative structure activity relationship, to combine the engineering control afforded by rational design with the screening flexibility of directed evolution. To the best of our knowledge, this approach has never been used within environmental engineering even though it could have numerous applications for remediation of priority contaminants in environmental media.

Materials and Methods

Materials

Polyethylene glycol (PEG), lithium acetate (LiAc), tris-HCl (pH 7.5, 1 M), ethylenediaminetetraacetic acid (EDTA), 2,2'-azino-bis(3-ethylbenzthiazoline-6-sulfonic acid) diammonium salt (ABTS), 17β -estradiol, and hydrogen peroxide (29.9%, ACS reagent grade) were from Sigma-Aldrich (St. Louis, MO). Plasmid pYEXS1-HRP (Morawski et al., 2000) was the kind gift of Dr. F.H. Arnold (California Institute of Technology). Plasmid Midiprep Kit was from QIAGEN (Valencia, CA). Primers were from Invitrogen (Carlsbad, CA). BL21(de3) electrocompetent *E. coli* cells were from the Marsh Laboratory, Department of Chemistry and Biological Chemistry, University of Michigan. QuikChange® XL Site-Directed Mutagenesis Kit and XL 10-Gold Ultracompetent *E. coli* cells were from Stratagene (La Jolla, CA). Carrier DNA was from Clontech (Mountain View, CA). Protease-deficient *S. cerevisiae* strain BJ5465 (*ade2-1*, *ura3-52*, *trp3-11*, *pep-his*, *pobl11*, *6R*, *can1-100*, *GAL*) was obtained from the American Type Culture Collection (ATCC; Manassas, VA).

1. QSAR-Assisted Mutation Design for Enhanced Target Degradation Rate

1.1 Preliminary Sensitivity Assessment—Sensitivity of binding distance to single mutations within the HRP active site was assessed using X-Score, software designed to facilitate prediction of structure-based ligand binding affinity during rational drug design (Wang and Wang, 2002). Output is expressed as negative logarithms of the enzyme-substrate dissociation constant (pK_D). Sensitivity assessment was undertaken via alanine scanning mutagenesis; i.e., systematic substitution of an alanine for each of seventeen amino acid residues included in a previously-validated computational model of the enzyme's catalytic

pocket. This is a standard method for determining which residues are critically important during protein-ligand interactions (Wells, 1991; Diller et al., 2010). Locations tested as part of the ALA scan included the following residues: ASN70, ASN175, ASN72, ASP247, GLU64, GLY69, LEU131, LEU138, LEU223, PHE41, PHE68, PHE142, PHE143, PHE179, PHE221, PRO139, and PRO141. Structural coordinates corresponding to mutation at each of these locations were generated from a model for HRP compound II that was originally based on full protein coordinates downloaded from the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) (Berglund et al., 2002). HyperChem Molecular Modeling System v. 7.1 (Hypercube, Inc.: Gainesville, Florida) was used to generate ten standardized (static) geometries of a single 17 β -estradiol molecule at different locations within the distal region of HRP's active site. These structures were subsequently converted from *.ml2 format to the *.mol2 format required by X-Score using commercial translation package Mol2Mol v. 5.4. X-Score was then used to assess van der Waals and hydrophobic interactions in wild type and ALA-mutant HRP.

1.2 Identifying Suitable Replacements for Sensitive Residue Locations—Various candidates were evaluated for substitution into locations previously occupied by active site residues to which 17 β -estradiol binding distance was deemed most sensitive. The anticipated effect of each mutation was assessed using a previously-validated molecular dynamics simulation protocol to compute 17 β -estradiol binding distances for each mutant (Colosi et al., 2006; Colosi et al., 2007b). Briefly, a single substrate molecule was aligned into a random location within the distal region of HRP's mutated active site. The OPLS molecular mechanics force field was then utilized in a Monte Carlo energy-minimization molecular dynamics binding simulation in which the ligand and two key docking residues (HIS42 and ARG38) were allowed to move freely, but the rest of the protein's active site was held fixed. (See Supporting Information for full simulation details.) Distances between 17 β -estradiol's phenolic proton and the imidazole- δ N on HIS42 responsible for electron transfer were recorded for each time step. These were synthesized into an energy-weighted average binding distance for $n > 50,000$ time steps, whereby distances corresponding to more probable lower-energy conformations were weighted more heavily than those corresponding to less probable higher-energy conformations. Mutations mediating a significant reduction in average 17 β -estradiol binding distance (μ_{HIS}) were considered promising candidates for further evaluation.

2. Empirical Validation of QSAR-Predicted Changes in Enzyme Activity

2.1 Expression of HRP Mutants in *S.cerevisiae*—Plasmid pYEXS1-HRP containing the gene for HRP-C, the Amp^r gene, and yeast selectable marker URA3 was transformed into BL21(de3) electrocompetent *E. coli* cells via electroporation. Cells were incubated for 1 hour at 37 °C with shaking. Aliquots were then transferred to plates of LB/amp medium and incubated overnight. Six individual colonies selected from these plates were used to inoculate 25-mL liquid cultures in LB/amp medium and returned to incubation overnight. A QIAGEN Plasmid Midi Kit was used to extract DNA according to the manufacturer's instructions. Subsequent sequencing confirmed presence and correct orientation of the pYEXS1-HRP plasmid. Sequencing primer was 5'-CGTAGTTTTTCAAGTCTTAG-3' (Morawski et al., 2000).

Aliquots containing 10 ng of purified, sequence-confirmed wild type plasmid were mutated using the QuikChange® XL Site-Directed Mutagenesis Kit according to the manufacturer's directions. Eight sets of primers corresponding to mutations of interest were designed using web-based program PrimerX (Lapid, 2003). Forward sequences are indicated in Supporting Information Table S1. Resulting plasmids were then transformed via heat-shock into XL10-Gold Ultracompetent *E. coli* cells, and 250- μ L aliquots of each transformed cell solution were spread in duplicate onto selective LB/amp agar plates. Following overnight incubation, six

colonies of each mutation were used to inoculate individual 25-mL cultures in the same selective media. These six cultures were incubated overnight, and then mutant plasmid was isolated and sequenced as above.

Plasmids corresponding to wild type HRP and eight different mutations were transformed into *S. cerevisiae* strain BJ5465. 10 μ L of transforming plasmid was added to a microcentrifuge tube also containing 100 μ g of denatured carrier DNA, 0.5 mL of PLATE solution (40% PEG, 0.1 M LiAc, 10 mM Tris-HCl, 1mM EDTA), and one colony of BJ5465. Tubes were vortexed briefly and incubated at room temperature for roughly 48 h. These were then immersed in a 42° C water bath for 15 min and spun at 10,000 rpm in a microcentrifuge. Supernatant was removed from each tube, and cells were resuspended in 200- μ L volumes of sterile DI. Resulting suspensions were spread onto plates of selective yeast nitrogen base medium without uracil (YNB/URA⁻) comprising 0.67% yeast nitrogen base without amino acids, 0.2% URA⁻ medium supplement, 2% glucose, and 0.50 mL/L of a stock trace metal solution (0.5 g/L CaCl₂·2H₂O, 0.2 g/L CoCl₂·6H₂O, 1.3 g/L CuCl₂·2H₂O, 34.5 g/L FeCl₃·6H₂O, 0.04 g H₃BO₃, 1.1 g/L MgCl₂·6H₂O, 1 g/L Na₂MoO₄·2H₂O, 0.7 g/L ZnCl₂·4H₂O). Plates were incubated overnight at 30° C before transfer to cold storage at 4° C for up to one month.

2.2 Kinetic Evaluation of Recombinant Enzyme (HRP*)—Three individual colonies were picked from selective YNB/URA- plates of *S. cerevisiae* expressing either wild type HRP or one of the eight mutants. These were used to inoculate three 5-mL pre-cultures in liquid YNB/URA- for 16 h at 30° C with shaking. Pre-cultures were then combined, transferred to Erlenmeyer flasks containing 135 mL of fresh YPDA nonselective medium (1% yeast extract, 2% peptone, 2% dextrose, 0.003% adenine), and returned to incubation for an additional 48 h. Cultures were then centrifuged at 1410g for 10 min to separate biomass. Supernatant containing extracellular HRP* was collected, dosed to 40% (m/v) with (NH₄)₂SO₄, and pumped onto a HiTrap Phenyl Fast Flow column (Low Sub, 1 mL; GE Healthcare, USA) at a flow rate of 0.75 mL/min. Absorbance was monitored continuously at 405 nm, the wavelength corresponding to maximal absorbance by the protein's heme prosthetic group. After 45 min, flow was switched to distilled deionized water (DDI), and HRP* was eluted from the column in ~0.5 mL of pure DDI (characteristic retention time, t_R = 2.0 min). This process was repeated four to six times, and collected fractions were pooled together. Absorbance at 405 nm (Abs_{405}) was then used to estimate HRP concentration. To minimize the effect of variation in HRP* activity among cultures evaluated on different days, this process was performed three times for each strain.

Initial rates of enzyme-mediated substrate removal were measured using a previously-reported method (Colosi et al., 2006; Colosi et al., 2007b) with slight modification. In brief, reactions were carried out at room temperature in 2 mL of phosphate buffer solution (PBS; 10 mM, pH = 7.0) using 7-mL Teflon-capped glass test tubes as completely mixed batch reactors (CMBRs). Initial concentrations (S_0) of 17 β -estradiol varied over the range 0.5–15 μ M, as prepared from a 15 μ M stock in PBS containing 5% methanol to enhance solubility. Phenol solutions were prepared from a 8000 μ M stock in PBS for evaluation of degradation over the range of S_0 = 50–6000 μ M. The volume of freshly-harvested HRP* required for each set of reactions was calculated based on Abs_{405} of the pooled sample being evaluated in that particular experiment. In general, enzyme dosage was about 0.001 U/mL (approximately 0.06 nM), where one unit of wild type HRP activity is defined as the quantity catalyzing oxidation of 1 μ mol ABTS per minute. Reaction in each CMBR was initiated upon addition of 150 μ M H₂O₂, an amount determined in preliminary tests as sufficient to saturate the wild type enzyme. Test tubes were then capped and shaken by hand for 20 s before 35 μ L of 1.0-N HCl was added to terminate the reaction. Duplicate experiments and a blank (equivalent volume buffer used in place of hydrogen peroxide) were performed for each initial substrate concentration. Immediately following each reaction, test tubes were centrifuged at 2205g for 20 min.

Supernatant from each CMBR was transferred to 2-mL glass vials for determination of pre-reaction (blank = S_0) and post-reaction (S_t) substrate concentrations using an Agilent 1100 series high-performance liquid chromatograph (HPLC) equipped with a Phenomenex Hypersil C18 column (250 × 2.0 mm, 5- μ m particle size), a UV-Diode Array Detector (DAD), and a fluorescence detector. Detection of 17 β -estradiol was achieved using the fluorescence detector with excitation at 239 nm and emission at 314 nm. Mobile phase was 40% HPLC-grade acetonitrile (ACN) and 60% DDI flowing at 0.6 mL/min. Retention time for 17 β -estradiol was t_R = 4.3 min. Detection of phenol was achieved using the UV-DAD set to 275 nm, and a mobile phase of 25% ACN + 75% DDI flowing at 0.45 mL/min. Retention time for phenol was t_R = 4.1 min.

Measured substrate concentrations were used to define initial reaction rate (r_0) according to $r_0 = (S_0 - S_t) / \Delta t$, where Δt was always 20 s. Values for r_0 were then plotted as a function of S_0 and fit to the classical Michaelis-Menten equation: $r_0 = r_{MAX} \times S_0 / (K_M + S_0)$; where r_{MAX} is the maximum rate of reaction and K_M is the substrate's Michaelis constant. By definition, the maximum reaction rate, $r_{MAX} = k_{CAT}[E]$, is achieved when substrate concentration is sufficiently large ($S_0 \gg K_M$) to saturate all available enzyme molecules $[E]$. Although $[E]$ is in actuality reduced over time as the enzyme becomes inactivated, it was assumed that the 20s reaction period was sufficiently short such that $[E]$ could always be approximated by $[E_0]$. This assumption makes it possible to define an additional parameter, the so-called "turnover rate", according to $k_{CAT} = r_{MAX} / [E_0]$, where E_0 was estimated to be approximately 0.06 nM on the basis of spectrophotometric measurements of a standard ABTS assay at λ = 405 nm.

Results and Discussion

1. QSAR-Assisted Mutation Design to Enhance Reactivity towards Target Substrate

1.1 Preliminary Sensitivity Assessment—Given the vast number of possible mutations to choose from for our re-design, and any protein design project in general, we began by evaluating the sensitivity of our target substrate's binding interaction to each residue within a previously-validated model of the enzyme active site. This was done computationally using the freely-available X-Score program. For simplicity, only portions of the enzyme with a documented role in substrate binding were incorporated into our computational model of the enzyme's active site. These included: the porphyrin ring, the heme iron and its double-bonded oxygen atom, twenty-two amino acids, two calcium atoms adjacent to the catalytic pocket, and a single calcium-associated water molecule (HOH311). Of the twenty-two residues included in the original active site model, three (HIS42, HIS170, and ARG38) were excluded from sensitivity assessment because they are critical to HRP functionality (Poulos and Kraut, 1980; Rodriguez-Lopez et al., 1996). Additionally, the nature of the sensitivity assessment technique, whereby each active site residue was systematically substituted by an alanine, made it impossible to evaluate ALA residues included in the original active site model (ALA71 and ALA140). X-Score was thus used to characterize van der Waals (VDW) and hydrophobic (HP) interactions between ALA-substituted mutants for the remaining seventeen amino acid locations using ten standardized conformations of the design substrate (17 β -estradiol). Figure 1 depicts resulting sensitivity measurements, wherein sensitivity is operationally defined as the average difference between wild type (WT) scores and ALA-mutant scores.

Mutations exhibiting large differences between the WT score and their ALA-scanning mutagenesis score were deemed most critical during interaction between HRP and the target ligand. Thus these were deemed most promising for further analysis. X-Score output was thus used to rank mutation candidates from most to least promising on the basis of sensitivity as follows: LEU138, PHE142, LEU131, PHE68, PHE179, GLU64, PHE41, ASN70, PHE143, PRO141, GLY69, ASP247, LEU223, PHE221, ASN72, and ASN175. This ranking was then cross-referenced against a list of previously-evaluated active site mutations (Supporting Info,

Table 1) to identify locations which should not be altered. Sensitivity rankings were then revised to reflect overall suitability for strategic substitution according to: LEU138, PHE142, LEU131, PHE68, PHE179, PHE41, LEU223, and ASN175.

1.2 Identifying Suitable Replacements for Sensitive Residue Locations—In

redesigning a protein catalyst, it's necessary to identify not only which locations should be changed, but also which amino acids should be substituted into these locations. The QSAR-assisted approach makes it possible to assess this computationally as follows. First, an array of replacements for each of the most sensitive locations was identified, including substitutes with both larger/smaller molecular volumes and increased/decreased hydrophobicity (as quantified using octanol-water partitioning coefficients, logP) relative to the original amino acid at each location. We then used a previously-validated molecular dynamics simulation protocol, in which the ligand and two key docking residues (HIS42 and ARG38) were allowed to move freely, to compute expected binding distances for each of these mutations. These computations were required because our QSAR model parameterizes enzyme-substrate interaction using average intermolecular binding distance; however, the exact nature of QSAR parameters used for redesign in other applications could vary depending on the enzyme/substrate system being redesigned.

Results of our mutation screening analyses are summarized in Figure 2, wherein, substituted locations are presented from left to right, using the same ranking (most seemingly suitable to least seemingly suitable) referenced above. Predicted binding distance for the wild type enzyme is indicated at far left and is highlighted by a dashed line of the same height to facilitate comparison with mutant predictions. As indicated in Figure 2, nine of the twenty-two simulated mutations exhibited no significant change in our QSAR parameter of interest, intermolecular binding distance, while an additional five mediated a dramatic increase. The remaining eight (L138M, F142Q, F142Y, L131P, F179A, F41A, F41L, and F41T) were associated with significant decrease. Interestingly, no mutation was able to achieve the seemingly optimal binding distance (7.49 Å) shared by smaller, more readily-degraded phenolic compounds in our previous QSAR analysis (Colosi et al., 2006). Thus, on the basis of the QSAR alone, it seems that HRP reactivity towards our target substrate (17β-estradiol) could be significantly improved but it's unlikely that this improvement will achieve the optimum reactivity exhibited HRP towards its native substrates.

Information summarized in Figure 2 offers several insights related to binding interaction between HRP's active site and the design substrate, 17β-estradiol. First, each of the eight mutations mediating significant reduction in binding distance involves replacement of a phenylalanine or leucine residue. Because these are two of the largest residue types included in the active site model, seven of the eight most promising mutations constitute a decrease in molecular volume. This observation is intuitively consistent with the notion of "opening up" the active site to render it more accommodating towards bulky, strongly-estrogenic substrates. Second, each of the eight mutations mediating significant reduction in 17β-estradiol binding distance also affects a decrease in hydrophobicity. Given the target substrate's highly aromatic nature, this is perhaps less intuitive than the relationship between binding distance and molecular volume; however, it's possible that decreased hydrophobicity in the area surrounding the heme may help to "push" the molecule towards the polar HIS42-N responsible for abstracting an electron from the substrate (Poulos and Kraut, 1980). The high degree of correlation ($r = 0.75$) between molecular volume and hydrophobicity makes it difficult to assess which factor drives reduction in binding distance, so it could be either alone or a combination of both.

2. Empirical Validation of QSAR-Predicted Changes in HRP Reactivity

2.1 Kinetic Evaluation of HRP Mutant Reactivity towards the Design Substrate

—To validate the notion of QSAR-assisted protein design, it was necessary to show that a model enzyme, HRP, could be made more reactive towards a target pollutant. Specifically, it was necessary to produce and evaluate selected HRP mutants, which on the basis of computational QSAR investigations, seemed most likely to exhibit improved performance. Thus, mutated versions of a plasmid bearing a synthetic gene for HRP were transformed into *S. cerevisiae* to produce recombinant wild type HRP and eight mutant HRP strains. These strains corresponded to the eight mutations mediating significant reduction in simulation-predicted binding distance for 17 β -estradiol (see Figure 2). Table 1 summarizes Michaelis-Menten kinetics parameters for reactions between each resulting type of HRP and the target substrate, 17 β -estradiol. As indicated in the table, two of the mutations (F41A and F142Y) exhibit statistically significant increases in $\ln(r_{\text{MAX}})$ (at $\alpha = 0.05$), and all but two mutations (F41L and F41T) yield statistically significant differences at ($\alpha = 0.20$). Moreover, simulation-predicted 17 β -estradiol binding distances exhibit inverse *linear* correlation with $\ln(r_{\text{MAX}})$ (Figure 3), consistent with the hypothesis that poor binding impedes HRP reactivity towards “bulky” substrates. Although the magnitude of $\ln(k_{\text{CAT}})$ increase is somewhat less dramatic than in other protein engineering studies to date ($2\text{--}3 \times$ versus $20\times$), our observations validate the legitimacy of QSAR-assisted protein design for targeting otherwise recalcitrant emerging contaminants. In particular, the linearity of Figure 3 demonstrates that the performances of the selected mutants exhibit the QSAR-predicted linear relationship.

2.2 Kinetic Evaluation of HRP Mutants Reactivity towards a Native Substrate—

Having evaluated the effect of QSAR-designed mutations on HRP reactivity towards the design substrate, it was also of interest to evaluate reactivity towards a substrate that had previously been degraded at its QSAR-predicted optimum rate. Specifically, it was hypothesized that “opening up” the active site to accommodate a bulkier substrate might have either of the following unintentional effects on reactivity towards smaller substrates: 1) increased ability to accommodate any substrate, regardless of its size, with resultant increase in all substrate reaction rates; or, 2) decreased ability to maintain critical points of contact with small substrate features, mediating reduction in small substrate r_{MAX} compared to the wild type. Thus, phenol was reacted with wild type HRP and each of the two most promising QSAR-designed mutants (F41A and F142Y). The resulting r_{MAX} values, as presented in the last two columns of Table 1, are consistent with the first hypothesized outcome identified above. That is, both mutations mediate dramatic increase in maximum phenol degradation rate compared to the wild type. What’s more, the increase in $\ln(r_{\text{MAX}})$ for phenol degradation is directly proportional to simulation-predicted binding distances ($R^2 = 1.00$) for each mutant strain. Thus it seems that smaller substrates also benefit from enhanced access to the reaction center in a manner consistent with the QSAR. Finally, in as much as the slope of each correlation encapsulates reaction rate sensitivity per unit change in intermolecular binding distance, it seems that phenol binding is more sensitive than 17 β -estradiol binding. These substrates exhibit slopes of $-0.54 \text{ \AA}^{-1}$ and $-0.19 \text{ \AA}^{-1}$, respectively. This is unexpected since 17 β -estradiol, rather than phenol, was the design substrate.

Conclusions

In demonstrating that our QSAR model, in conjunction with simple computational screening, can help identify which mutations out of a seemingly infinite pool are most likely to increase enzyme reactivity towards a target substrate, this work demonstrates the validity of a novel “QSAR-assisted” approach to protein catalyst design. As proof of concept, this work describes one QSAR model for one model enzyme (HRP) and one target substrate; however, this approach could be deployed for other enzymes of interest. Although the magnitudes of

increased HRP reactivity exhibited in this study are somewhat modest compared to other investigations, we hope that this technique's unique advantages compared to conventional protein engineering via rational design or directed evolution, will galvanize future work in this area. Thus, two principal advantages should be emphasized. First, QSAR-assisted design leverages mechanistic understanding of factors affecting enzyme reactivity towards a class of substrates rather than just one single substrate. This is helpful in the design of environmental catalysts since there are so many large classes of priority contaminants. Second, use of an overarching QSAR makes it possible to combine some benefits of rational design and directed evolution. In particular, the QSAR-assisted approach retains the fundamental control of rational design and the rapid screening (albeit computational rather than experimental) flexibility of directed evolution. Ultimately, this investigation opens the door to more strategic techniques for advanced purification of water containing emerging contaminants.

Future work will assess the feasibility of designing waste-treatment catalysts on the basis of QSAR relationships that incorporate both maximum reaction rate and substrate specificity. Additionally, other classes of enzymes will be subjected to similar analysis now that the principle of QSAR-assisted protein design has been validated as means to enhance reactivity towards a target contaminant.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was financed by Research Grant P42ES04911-14 from the National Institutes for Environmental and Health Sciences. LMC is thankful for support of a National Science Foundation Graduate Research Fellowship during the time these experiments were performed. The authors also wish to acknowledge significant experimental assistance from Mrs. Christel C. Fox and Professor E. Neil G. Marsh.

References

- Ang EL, Zhao H, Obbard JP. Recent advances in the bioremediation of persistent organic pollutants via biomolecular engineering. *Enzyme Microb Technol* 2005;37:487–496.
- Arnold FH, Volkov AA. Directed evolution of biocatalysts. *Curr. Opin. Chem. Biol* 1999;3:54–59. [PubMed: 10021399]
- Auriol M, Filali-Meknassi Y, Tyagi RD, Adams CD. Oxidation of natural and synthetic hormones by the horseradish peroxidase enzyme in wastewater. *Chemosphere* 2007;68:1830–1837. [PubMed: 17498772]
- Berglund GI, Carlsson GH, Smith AT, Szoke H, Henriksen A, Hajdu J. The catalytic pathway of horseradish peroxidase at high resolution. *Nature* 2002;417:463–478. [PubMed: 12024218]
- Carmichael AB, Wong LL. Protein engineering of *Bacillus megaterium* CYP102-the oxidation of polycyclic aromatic hydrocarbons. *Eur. J. Biochem* 2001;268:3117–3125. [PubMed: 11358532]
- Caza N, Bewtra JK, Biswas N, Taylor KE. Removal of phenolic compounds from synthetic wastewater using soybean peroxidase. *Water Res* 1999;33:3012–3018.
- Cho CMH, Mulchandani A, Chen W. Functional analysis of organophosphorus hydrolyase variants with high degradation activity towards organophosphate pesticides. *Protein Eng. Des. Sel* 2006;19:99–105. [PubMed: 16423845]
- Cho CMH, Mulchandani A, Chen W. Bacterial cell surface display of organophosphorus hydrolyase for selective screening of improved hydrolysis of organophosphate nerve agents. *Appl. Environ. Microbiol* 2002;68:2026–2030.
- Colosi LM, Burlingame DJ, Huang Q, Weber WJ Jr. Peroxidase-mediated degradation of a PCB using natural organic matter as sole co-substrate. *Environ. Sci. Technol* 2007a;41:891–896.

- Colosi LM, Huang Q, Weber WJ Jr. QSAR-based quantification of the impacts of enzyme-substrate binding on rates of peroxidase-mediated reactions of estrogenic phenolic chemicals. *J. Am. Chem. Soc* 2006;128:4041–4047. [PubMed: 16551113]
- Colosi LM, Huang Q, Weber WJ Jr. Validation of a two-parameter QSAR as a legitimate tool for rational re-design of horseradish peroxidase. *Biotechnol. Bioeng* 2007b;98:295–299. [PubMed: 17657769]
- Colosi LM, Pinto RA, Huang Q, Weber WJ Jr. Peroxidase-mediated degradation of perfluorooctanic acid. *Environ. Toxicol. Chem* 2009;28:264–271. [PubMed: 18808217]
- Diller DJ, Humblet C, Zhang X, Westerhoff LM. Computational alanine scanning with linear scaling semiempirical quantum mechanical methods. *Proteins* 2010;78:2329–2337. [PubMed: 20544968]
- DiSioudi BD, Miller CE, Lai K, Grimsley JK, Wild JR. Rational design of organophosphorus hydrolyase for altered substrate specificities. *Protein Eng. Des. Sel* 1999;19:99–105.
- Fang H, Tong W, Shi LM, Blair L, Perkins R, Branham W, Hass BS, Xie Q, Dial SL, Moland CL, Sheehan DM. Structure-activity relationships for a large diverse set of natural, synthetic, and environmental estrogens. *Chem. Res. Toxicol* 2001;14:280–294. [PubMed: 11258977]
- Gianfreda L, Rao MA. Potential of extra cellular enzymes in remediation of polluted soils: a review. *Enzyme Microb. Technol* 2004;35:339–354.
- Harford-Cross CF, Carmichael AB, Allan FK, England PA, Rouch DA, Wong LL. Protein engineering of cytochrome P450 (cam) (CYP101) for the oxidation of polycyclic aromatic hydrocarbons. *Prot. Eng* 2000;13:121–128.
- Johnson AC, Sumpter JP. Removal of endocrine disrupting compounds in activated sludge treatment works. *Environ. Sci. Technol* 2001;35:4697–4703. [PubMed: 11775141]
- Joo H, Lin ZL, Arnold FH. Laboratory evolution of peroxide-mediated cytochrome P450 hydroxylation. *Nature* 1999;399:670–673. [PubMed: 10385118]
- Lapid C. PrimerX - Automated design of mutagenic primers for site-directed mutagenesis. 2003
- Morawski B, Lin Z, Cirino P, Joo H, Bandara G, Arnold FH. Functional expression of HRP in *Saccharomyces cerevisiae* and *Pichia pastoris*. *Protein Eng* 2000;13:377–384. [PubMed: 10835112]
- Nishihara T, Nishikawa J, Kanayama T, Dakeyama F, Saito K, Imagawa M, Takatori S, Kitagawa Y, Hori S, Utsumi H. Estrogenic activities of 517 chemicals by yeast two-hybrid assay. *J. Health Sci* 2000;46:282–298.
- Poulos TL, Kraut J. The stereochemistry of peroxidase catalysis. *J. Biol. Chem* 1980;255:8199–8205. [PubMed: 6251047]
- Purdum CE, Hardiman PA, Bye WJ, Eno NC, Tyler CR, Sumpter JP. Estrogenic effects of effluents from sewage treatment works. *Chem. Ecol* 1994;8:275–285.
- Rodriguez-Lopez JN, Smith AT, Thorneley RNF. Role of arginine in horseradish peroxidase. *J. Biol. Chem* 1996;271:4023–4030. [PubMed: 8626735]
- Salazar O, Cirino PC, Arnold FH. Thermostabilization of a cytochrome P450 peroxygenase. *Chembiochem* 2003;4:891–893. [PubMed: 12964165]
- Wang R, Lai L, Wang S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput. Aided Molec. Des* 2002;16:11–26. [PubMed: 12197663]
- Watkins LM, Mahoney HJ, McCulloch JK, Raushel FM. Augmented hydrolysis of diisopropyl fluorophosphate in engineered mutants of phosphotriesterase. *J. Biol. Chem* 1997;272:25596–25601. [PubMed: 9325279]
- Wells JA. Systematic mutational analyses of protein-protein interfaces. *Meth. Enzymol* 1991;202:390–411. [PubMed: 1723781]

Figure 1.

Sensitivity of QSAR substrate binding distance to ALA-substitution at sixteen active site locations as computed using X-Score. WT indicates score for wild type HRP. VDW and HP indicate portions of the score associated with Van der Waals and hydrophobic interactions, respectively. Error bars represent 90% confidence intervals for $n = 10$ standardized substrate orientations. Alteration at PRO139 made the X-Score model computationally unstable so data are not shown.

Figure 2.

Results from computational screening of possible mutations to enhance enzyme reactivity towards target substrate (17 β -estradiol), specifically, intermolecular binding distances for selected active site mutations. Error bars represent 95% confidence intervals for $n = 50,000$ enzyme-substrate configurations. Dashed line indicates magnitude of wild type (WT) binding distance, 12.40 Å. Asterisks (*) indicate mutations mediating significant reduction in QSAR parameter of interest (binding distance) compared to WT.

Figure 3.

Measured maximum initial reaction rates for transformation of 17 β -estradiol by recombinant wild type (WT) HRP and eight mutants. Error bars represent ± 1 SE for best-fit Michaelis-Menten parameterizations; asymmetry reflects use of the Ln- transformation.

Table 1

Simulation-predicted distances and Michaelis-Menten kinetic parameters for reaction between recombinant HRP strain and two substrates.

HRP Strain	17 β -Estradiol			Phenol		
	Binding Distance (Å)	r _{MAX} (μ M/s)	Ln[K _{CAT} /K _M]	Binding Distance (Å)	r _{MAX} (μ M/s)	Ln[K _{CAT} /K _M]
F41A	9.03	0.056 $\pm$ 0.016	5.6	5.53	43.6 $\pm$ 8.30	3.9
F142Y	9.58	0.038 $\pm$ 0.005	5.5	5.66	39.6 $\pm$ 49.1	3.9
F142Q	10.25	0.038 $\pm$ 0.009	4.8			
L138M	10.47	0.034 $\pm$ 0.006	5.9			
F179A	10.52	0.040 $\pm$ 0.010	4.9			
F41T	10.80	0.032 $\pm$ 0.005	4.6			
F41L	11.02	0.033 $\pm$ 0.006	5.2			
L131P	11.04	0.035 $\pm$ 0.007	5.5			
Wild Type	12.40	0.027 $\pm$ 0.004	5.1	7.40	15.6 $\pm$ 19.1	3.4