

**Consensus Mutagenesis Reveals that Non-Helical Regions Influence Thermal  
Stability of Horseradish Peroxidase.**

Barry J. Ryan<sup>1</sup>, Mary J. O'Connell and Ciarán Ó'Fágáin<sup>1\*</sup>.

School of Biotechnology and <sup>1</sup>National Centre for Sensors Research, Dublin City  
University, Dublin 9, Ireland.

\* corresponding author: Ciarán Ó'Fágáin (Ciaran.Fagan@dcu.ie).

Phone: +353-1-700 5288

Fax: +353-1-700 5412

**Key Words:** Ancestral Protein; Consensus Concept; Protein Stabilisation; Recombinant  
Horseradish Peroxidase.

**Short Title:** Consensus mutagenesis of horseradish peroxidase

## **Summary**

The enzyme horseradish peroxidase has many uses in biotechnology but a stabilized derivative would have even wider applicability. To enhance thermal stability, we applied consensus mutagenesis (used successfully with other proteins) to recombinant horseradish peroxidase and generated five single-site mutants. Unexpectedly, these mutations had greater effects on steady-state kinetics than on thermal stability. Only two mutants (T102A, T110V) marginally exceeded the wild type's thermal stability (4% and 10% gain in half-life at 50°C respectively); the others (Q106R, Q107D, I180F) were less stable than wild type. Stability of a five-fold combination mutant matched that of Q106R, the least-stable single mutant. These results were perplexing: the Class III plant peroxidases display wide differences in thermal stability, yet the consensus mutations failed to reflect these natural variations. We examined the sequence content of Class III peroxidases to determine if there are identifiable molecular reasons for the stability differences observed. Bioinformatic analysis validated our choice of sites and mutations and generated an archetypal peroxidase sequence for comparison with extant sequences. It seems that both genetic variation and differences in protein stability are confined to non-helical regions due to the presence of a highly conserved alpha-helical structural scaffold in these enzymes.

**Abbreviations:** ABTS, 2,2'-azino-bis(3-ethylbenzthioline-6-sulfonic acid;  $\delta$ -ALA, delta aminolevulinic acid; DMF, dimethylformamide; DMSO, dimethylsulfoxide; HRP, horseradish peroxidase isoenzyme C; HRPA2, horseradish peroxidase acidic isoenzyme A2; GnCl, Guanidine Hydrochloride; IPMDH, 3-isopropylmalate dehydrogenase; LB, Luria-Bertani medium; MeOH, methanol; ML, maximum likelihood; MRCA, most recent common ancestor; NCBI, National Centre for Bioinformatic Information; PAML, phylogenetic analysis using maximum likelihood; PEG, polyethylene glycol; RZ, reinheitszahl (purity number;  $A_{403}/A_{280}$ ); SBP, soybean peroxidase;  $t_{1/2app}$ , apparent half-life; v/v, volume per volume; w/v, weight per volume.

## Introduction

The peroxidases (E.C. 1.11.17) are a ubiquitous subset of enzymes found throughout the animal and plant kingdoms [1]. The enzyme from horseradish roots (*Armoracia rusticana*; HRP) is the most widely studied example, due mainly to its many diverse uses in biotechnology [2]. Although moderately stable, the availability of a stabilized form of HRP would increase its applicability still further. Previous stabilisation studies have focused on the plant-derived protein, with several reports describing chemical procedures such as crosslinking [3-6], surface modification [7-9], attachment of PEG [10] and modification of carbohydrate residues [11]. Immobilisation of HRP [12, 13] and addition of stabilising reagents [14, 15] have also led to enhanced stability. To date, random mutagenesis has been used to genetically alter recombinant HRP stability [16, 17] while yeast cell surface display has been used to select recombinant HRP variants with altered enantioselectivities [18]. There have been no reported attempts to stabilise recombinant HRP via site-directed mutagenesis.

It is not necessary to examine a protein's three-dimensional structure in order to identify stabilizing substitutions. Using sequence information alone, Steipe et al. [19] successfully predicted stabilizing mutations in a  $V_H$  antibody domain with >60% frequency. This "Consensus" alignment procedure allows the identification of key stabilising residues in a protein structure [20]. The approach postulates that conserved residues in the sequence alignments of related proteins are more stabilising than non-conserved residues [21] and that a set of amino acid sequences of homologous, mesophilic enzymes contains sufficient information to allow rapid design of a thermostabilised, fully functional enzyme [22]. The chances of a deleterious mutation are reduced, since the replacement amino acid has already proven its evolutionary fitness at that position.

A consensus peroxidase protein sequence was developed “*in silico*” from aligned sequences of Class III secretory plant peroxidases downloaded from NCBI utilising custom-built bioinformatic software that scores the most frequent amino acid in a particular position [23]. Using this consensus sequence, five HRP residues were mutated to the corresponding consensus residue. The mutant proteins were expressed, purified and characterised in terms of their stabilities and steady-state ABTS kinetics.

Only two mutants registered marginal stability gains over wild type recombinant HRP, in sharp contrast to previous consensus studies with other proteins [19-22, 24-26]. This unexpected outcome prompted us to undertake an evolutionary analysis of the Class III peroxidases and to generate a hypothetical archetypal peroxidase sequence. Comparison of extant sequences with our hypothetical archetypal peroxidase reveals an interesting insight into plant peroxidase structure, function and evolution.

### **Materials and Methods.**

*Materials.* All reagents were purchased from Sigma Aldrich and were of analytical grade or higher. The pQE60 vector was purchased from Qiagen (Valencia, CA); XL 10 Gold cells and QuickChange<sup>TM</sup> Mutagenesis Kit were purchased from Stratagene (La Jolla, CA). The HRP gene was a generous gift from Prof. Frances H. Arnold (Caltech, CA, USA).

*Cloning.* Based on refs. [27, 28], the HRP gene was directionally cloned into the pQE60 vector as a fusion with the pectate lyase (PelB) leader sequence [29] (preceding the HRP’s N-terminus) and a hexa-histidine purification tag (at the C-terminus), to generate plasmid pBR\_I [30].

*Bacterial Strains and Plasmids.* *E.coli* XL 10 Gold was used as host strain to express the HRP protein. The plasmid pBR\_I, carrying the HRP gene coding for the HRP fusion protein, was used for expression and site directed mutagenesis.

*Recombinant DNA Techniques.* All DNA manipulations were carried out by standard techniques [31]. Site directed mutagenesis was carried out as described in ref. [32] utilising the QuickChange™ method. Mutant primers were supplied by MWG-Biotech (Germany). Mutants were confirmed by commercial di-deoxy sequencing (Fusion Antibodies, Belfast, Northern Ireland).

*Expression and Purification.* A single cell transformed with pBR\_I (or mutant derivative) was grown in 10mL LB medium containing 100 µg/mL ampicillin and 2% w/v glucose until the OD<sub>600nm</sub> reached 0.4; the cells were removed via centrifugation at 2,000 × g for 5 min and resuspended in fresh LB (500 mL) supplemented with 100 µg/mL ampicillin, 1mM δ-ALA and 2mM CaCl<sub>2</sub>. The cells were then allowed to grow at 30°C, 220 rpm for 16 h. Following overnight expression, the cells were centrifuged at 2,000 × g for 5 min and the supernatant was treated with 50% w/v ammonium sulphate for 2 h at room temperature. The cells were periplasmically lysed [33] and the periplasmic contents were similarly treated with 50% w/v (with respect to the initial supernatant volume) ammonium sulphate. Proteins precipitated by ammonium sulphate from both the culture supernatant and the periplasmic preparation were collected via centrifugation, resuspended in 50mM phosphate buffer pH 7.5, pooled and dialysed versus the same buffer overnight at 4°C. Sodium chloride (1M) and GnCl (200mM) were added to the dialysed fractions (10 mL total volume), and these latter were purified via nickel affinity chromatography at room temperature. Sodium acetate (25mM, pH 4.5) was utilised to elute the bound HRP. The eluted HRP was again dialysed versus 50mM phosphate buffer pH 7.5 overnight at 4°C, after which the protein was concentrated (Amicon-Plus 20 concentrator tubes; 2 mL

final volume), filter sterilised and stored at 4°C. These procedures led to typical expression values of 0.086 mg HRP per litre of culture medium. Purified HRP (specific activity 0.58  $\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$ , RZ value 1.1) gave a single band on a 12% polyacrylamide gel (not shown).

*Enzyme Assay and Characterisation.* The stability parameters of recombinant HRP and mutant variants were determined as described for plant HRP [3, 9] except that thermoinactivation time courses used 50°C. Samples were removed periodically onto ice and their residual activities determined upon re-warming to room temperature; this procedure gives apparent half-life,  $t_{1/2\text{app}}$ ). A constant protein concentration of 0.1 mg/mL was used for all thermoinactivations to control for possible effects of protein concentration on stability. The substrate ABTS (2,2'-azino-bis(3-ethylbenzthioline-6-sulfonic acid)) gives steady-state kinetics, permitting estimation of the apparent kinetic parameters,  $V_m/E$  and  $K'_m$  [34, 35].

*Generation of Consensus and Ancestral Sequences.* Approach 1: The *peroxidase consensus sequence* was generated *in silico* by the 'Protein Parser' software [23]. One hundred fully confirmed peroxidase protein sequences were downloaded from the NCBI homepage [<http://www.ncbi.nlm.nih.gov/>], using the following search terms “((((((Peroxidase) AND (plant))) NOT (precursor)) NOT (putative)) NOT (segment)) NOT (catalase)”. The sequences were then aligned via the Clustal W alignment package [36], using the default parameters. The alignment was saved as a “.aln” file and subsequently processed by Protein Parser [23] with a tolerance level of “50% Consensus” (i.e.  $\geq 50\%$  frequency of occurrence of a particular amino acid) to generate the consensus sequence.

Approach 2: Using the fully resolved phylogenetic tree for Class III plant peroxidases (with bootstrap support values  $> 70\%$  reported on all nodes; [37]), a single representative of each clade was selected at random and the corresponding multiple sequence alignment

was generated using the default parameters in Clustal W [36]. The alignment consists of 24 taxa and is 455 positions in length, see Figure 1(a). This multiple sequence alignment was subsequently parsed by ProteinParser producing a new consensus sequence. This second consensus sequence is free from any sampling bias or over-representation of sequences or clades, and serves as a control for the consensus sequence resulting from Approach 1.

Approach 3: To test our consensus method used in this analysis, we reconstructed all ancestral nodes on the reduced phylogenetic tree of the peroxidases. This approach determines the full-length peroxidase sequence that existed in the Most Recent Common Ancestor (MRCA) of the extant enzymes soybean peroxidase (SBP) and HRP. This sequence should concur with the features that the consensus approach determines. Using the maximum likelihood (ML) method implemented in Paml [38, 39] for marginal reconstruction of ancestral sequences, all ancestral nodes on the phylogenetic tree were determined (see Figure 1(a)). More specifically the extinct protein reconstruction was achieved using the aaml program (codeml.c with “seqtype=2”) in the PAML package [39] allowing for a gamma distribution model of rates across sites. This applies Bayesian statistics to define the most probable character state for each ancestral character over the entire length of the sequence. This statistical approach has previously been shown to yield very high confidence intervals for reconstructed states (91-99% accuracy for lysozyme c sequences; [38]). The ancestral sequence reconstruction is available in the supplementary information. The full length MRCA sequence referred to as the “Archetype”, (see Figure 1(a) and 2), was then compared with the consensus sequences from Approaches 1 and 2. The conserved, and most likely critical, positions in peroxidase enzymes are evident in Figure 2.

## Results.

Wild type recombinant HRP was less stable than the plant enzyme, probably due to glycosylation of the latter [40]. Over a 10-min incubation period, the half-inactivation temperatures for recombinant and plant HRP were 50°C and 55°C respectively. His-tagged and non His-tagged wildtype recombinant HRP displayed identical thermal stabilities.

Wild type recombinant HRP differed from the Approach 1 consensus sequence at 5 positions (single-letter amino acid code, wild type residues first: T102A, Q106R, Q107D, T110V, I180F; see Figure 2) and each position was mutated within the pBR\_I plasmid to the corresponding consensus amino acid, whilst maintaining the leader sequence and purification tag.

All HRP thermoinactivations fitted to a first-order decay ( $r^2 > 0.93$  in all cases) up to 10 min at a constant 50°C, allowing estimation of a first-order  $k$ -value and, hence, an *apparent* half-life ( $t_{1/2}$ ; see Discussion). Only T102A and T110V displayed modest increases in  $t_{1/2}$  (4% and 10% respectively); the other mutants were less stable than wild type (Table 1). The five-fold combination mutant showed very poor thermal stability, close to that of Q106R, the least-stable single mutant (Table 1).

Regarding HRP steady-state kinetics with ABTS, our values of 482 s<sup>-1</sup> for  $V_m/E$  and 0.093 mM for  $K'_m$  (Table I) compare with literature values of 810 s<sup>-1</sup>, 0.27 mM and 670 s<sup>-1</sup>, 0.18 mM for plant and recombinant HRP (both RZ >3) respectively [33], obtained under similar conditions. (Measurement of individual rate constants has not yet been undertaken.) Only marginal differences in  $V_m/E$  (+/- 26%) are noted for mutants T102A, Q106R and T110V, while Q107D is indistinguishable from wild type (as is the combination mutant; Table I). These four mutations lie at the back of the substrate-binding site. I180F, located in a proximal helix, also has  $V_m/E$  indistinguishable from wild type. Effects on  $K'_m$  are more marked, with increases of 9.5-fold noted for Q107D, 5.8-fold for Q106R, 3.3-fold for

I180F, 2.2-fold for T102A and 1.9-fold for T110V (Table 1).  $K'_m$  of the combination mutant is two-thirds that of wild type, the only instance of a  $K'_m$  value lower than the wild type value (Table 1). These differences are perhaps due more to the location of the mutations (close to, but not in, the substrate binding site; [41]) than to the nature of the substitution. Changes observed with the large ABTS substrate, however, may not reflect these mutants' performance with alternative reducing substrates; we have not yet tested any others.

### **Discussion.**

The phylogenetic relationships within the class III plant peroxidase family have been fully resolved [37]. Soybean peroxidase (see Figure 1, Group H) and HRP isoenzymes C (see Figure 1, Group A1) and A2 (see Figure 1, Group A2) are closely related in this phylogeny, yet their stabilities differ markedly. Relative clade positions of HRPC, HRPA2 and SBP versus our reconstructed “Archetype” sequence are shown in Figure 1. SBP (69% sequence similarity to HRPC) is notably more thermostable than HRPC [42, 43] but HRPA2 (68% similar to HRPC) is much less so (after 10min at 45°C, HRPA2 (Biozyme HRP-5) retained only 36% initial activity, while HRPC and SBP (both Sigma) each retained >90% activity; A.-M. O'Brien, unpublished work in this laboratory). Clearly, closely-related peroxidases show notable differences in thermal stability, yet our consensus mutants failed to reflect these differences. This prompted us to analyse peroxidase sequences and phylogeny by complementary methods and to generate a hypothetical archetypal peroxidase sequence. Comparison of HRP, HRPA2 and SBP sequences with our hypothetical archetypal peroxidase (see supplementary material) reveals an interesting insight into plant peroxidase structure, function and evolution.

We examined kinetic stability (persistence of catalytic activity) rather than folding stability because HRP unfolding is not a two-stage process [44]. Unlike previous successful consensus studies with other proteins [19-22, 24-26], none of the five substitutions yielded a significant gain in HRP thermal stability. Indeed, our consensus mutations had a greater influence on  $K'_m$  for ABTS than on thermal stability (Table 1). (Note, however, that our inactivation rate constants ( $k$  values) were measured at a single elevated temperature and at uniform protein concentration. Possible temperature- and concentration-dependences of the various  $k$  values have not yet been explored by thermoinactivations performed at different temperatures and/or protein concentrations. Hence, our reported half-lives should be taken as *apparent* values. In addition, our steady-state kinetic experiments, undertaken at constant  $H_2O_2$  concentration, yielded values of  $V_m/E$  and  $K'_m$  for ABTS only. An alternative steady-state methodology [45], where  $H_2O_2$  concentrations are also varied, can give greater insights by permitting estimation of the rate constants  $k_1$ ,  $k_2$ ,  $k_3$  and  $k_4$ . (This last is a feature of recombinant HRP but does not occur in plant HRP [34])).

In previous studies, six out of ten [19] and six out of twelve [46] consensus mutations were thermostabilizing; it has also been noted that thermostabilization occurs in about 33% of the total consensus mutants generated [22]. Our definition of ‘consensus’ ( $\geq 50\%$  occurrence of a particular residue at a given site) was not unduly restrictive: five out of six stabilizing consensus substitutions in an antibody  $V_H$  domain occurred at frequencies much greater than 50% [19]. The apparent failure of the consensus approach to yield a thermostabilized HRP (at least under our experimental conditions) led us to examine the sequence content of Class III peroxidases. We wished to discern possible molecular reasons for (i) the observed thermal stability differences among these enzymes and (ii) the disappointing outcome of our experiments.

We located consensus mutations by counting amino acid frequencies at each position of a Clustal alignment of 100 plant peroxidase sequences [37] using our Protein Parser programme [23] (Approach 1). We validated this approach by two alternative methodologies, Approaches 2 and 3, described above in Methods. Approach 3 provided us with complete sequences of the ancestral enzymes that gave rise to the extant enzymes. The complete phylogenetic tree for over 70 Class III peroxidase sequences was previously constructed [36] but no ancestral sequence was disclosed in that report. We compared our consensus sequences from Approaches 1 and 2 with the ancestral sequence (Approach 3) of the most ancient reconstructed peroxidase (the first or “archetype” peroxidase, which pre-dates all other peroxidases) and found 85% agreement across the alignment.

All three approaches identified T102A, Q106R and Q107D. Both Protein Parser (Approach 1) and ancestral protein generation (Approach 3) revealed T110V and I180F. These correspondences confirm that our initial consensus-identification approach was valid.

Alignment of the relevant sequences reveals that HRPC consensus substitution T110V actually occurs in the more thermostable SBP but, curiously, the T102A and I180F substitutions occur in both SBP *and* in the less thermostable HRPA2 (see supplementary material). These coincidences suggest that the consensus approach may not be a fruitful one for plant peroxidases, irrespective of its success with other proteins.

In previous reports of consensus protein thermostabilization, the various conserved columns are distributed evenly throughout the entire sequence. For example, Lehmann and Wyss [20] produced a consensus sequence for a fungal phytase consisting of single isolated columns of alignment. Four of our five consensus positions lie within a very short, nine-residue cluster spanning T102 - T110 in helix D (see Figure 2). Only I180 occurs outside this region (immediately before helix F'; see Figure 2). Such grouping is not

unique to peroxidases, however. Using a phylogenetic tree, Watanabe et al. [46] prepared 12 ancestral mutants of dimeric 3-isopropylmalate dehydrogenase (IPMDH), six of which were thermostabilized. Although the full set of 12 mutations was distributed throughout the polypeptide sequence, four clustered within positions 53-61 while another three occurred within positions 179-184. They also noted that the consensus and ancestral mutation approaches frequently indicated the same residue substitutions, as found in the present study. In the case of IPMDH, however, it appears that the ancestral residues are thermostabilizing because they occur in a hyperthermophilic common ancestor and not because they are conserved [46]. Mammalian antibodies, in contrast, are much more recent in origin and any stability gains for these proteins arising from consensus mutation (e.g. [19]) are unlikely to arise from the antiquity of the substituting residue [46]. It is possible that the amino acid sequences comprising the helix D region of plant peroxidases have already maximized thermostabilizing interactions for this region of the protein and that more global stability determinants, not revealed by consensus alignments, occur elsewhere in the molecule.

Thermal stability differences between homologous proteins may be due to a very few naturally occurring sequence variations ([47] and references within). Magliery and Regan [48] have developed a more advanced model which addresses some of the inadequacies of the basic consensus approach [22]. The underlying principle of their improved consensus method defines the probability of occurrence of a particular residue at a specific position compared to a reference state; i.e. it asks “how free is a position to vary, or change to another amino acid?” This method is potentially a very powerful test of covariance: the free energies of pairs of positions can be compared to determine if substitutions observed at one position in the sequence compensate for those observed at the other position [48]. In the present analysis, however, we wished merely to identify single

positions that vary between peroxidase groups, as these positions are candidates for conferring stability.

Gajhede et al. [41] had earlier noted that sequence variability within the Class III peroxidases was confined to certain "hot spots" in the molecule, including the substrate access channel and part of the region surrounding the interaction site of the aromatic donor molecule. From previous studies on peroxidases within *A. thaliana*, these enzymes share on average 75% identity in their primary structure [49]. This indicates that within the *A. thaliana* clade there is strong selective pressure to retain the structural conformation (HRP A2 is 95% similar to *A. thaliana* A2 peroxidase; [50]). We suggest that this is the case throughout the evolutionary history of the peroxidases *in toto*, as we observe that the ancestral peroxidase generated in this study also follows the same structural form. The Class III peroxidases have evolved to retain their structural helices and active site, suggesting that structure and function are tightly associated in this family: see Figure 1(b) and Figure 2. Retention of such structures suggests that these enzymes have a very ancient function consistent with their modern day one. There is some slight sliding of helices throughout evolution but this is acceptable over longer periods of time [51]. On the other hand, peripheral regions of a protein outside the helices are free to change their sequence (and folding pattern) entirely [51].

The present consensus residues clearly have little influence on peroxidase thermal stability. The dominance of a highly conserved alpha-helical structural scaffold suggests that the consensus approach is not a suitable protein stabilisation strategy for plant peroxidases, despite its success with other proteins. The determinants of differential thermal stabilities seen among plant peroxidases (e.g. HRP A2, HRP C and SBP) may lie in the unstructured loops linking the conserved helices. Unfortunately, these non-helical loops show a low incidence of consensus amino acids, so rational or semi-rational design

of stabilizing mutations within these potentially highly variable loops remains a challenging task. The ancestral protein we have reconstructed will serve as our starting point for this analysis.

Interestingly, a recent paper proposes that Met 37 of the SBP holoprotein is a key determinant of that enzyme's enhanced stability: this Met residue forms a direct van der Waals interaction with the C8 vinyl substituent of the heme prosthetic group which does not occur in holo-HRP [52]. (The stabilities of both heme-free apoproteins are very similar.)

Finally, although our consensus approach did not yield thermostabilized HRP variants, mutant T110V displayed significantly-enhanced resistance to inactivation by excess hydrogen peroxide [53].

#### **Acknowledgments**

We thank the Irish Research Council for Science, Engineering and Technology (Embark Initiative postgraduate scholarship to BJR) and Dublin City University (Postgraduate Accommodation Award to BJR, Albert College Award to CÓ'F). We are grateful to Prof FH Arnold and California Institute of Technology for the generous gift of a recombinant horseradish peroxidase plasmid and to Dr JO McInerney (National University of Ireland, Maynooth) for advice and use of the Bioinformatics and Pharmacogenomics Lab IBM cluster for the analysis of phylogeny and ancestral protein reconstruction. The National Centre for Sensors Research was established under the Higher Education Authority's Programme for Research in Third Level Institutions.

#### **References.**

1. Veitch, N.C., *Horseradish peroxidase: a modern view of a classic enzyme*. *Phytochemistry*, 2004. **65**(3): p. 249-59.
2. Ryan, B.J., N. Carolan, and C. Ó'Fágáin, *Horseradish and soybean peroxidases: comparable tools for alternative niches?* *Trends Biotechnol*, 2006. **24**(8): p. 355-63.
3. Ryan, O., M.R. Smyth, and C. Ó'Fágáin, *Thermostabilized chemical derivatives of horseradish peroxidase*. *Enzyme Microb Technol*, 1994. **16**(6): p. 501-5.
4. Miland, E., M. Smyth, and C. Ó'Fágáin, *Modification of horseradish peroxidase with bifunctional N-hydroxysuccinimide esters: effects on molecular stability*. *Enzyme and Microbial Technology*, 1996. **19**: p. 242-249.

5. O'Brien, A.M., Ó'Fágáin C, Nielsen PF, Welinder KG., *Location of crosslinks in chemically stabilized horseradish peroxidase: implications for design of crosslinks.* Biotechnol Bioeng, 2001. **76**(4): p. 277-84.
6. Gonera, A., P. Mischnick, and H. Ukeda, *The application of O-aminopropyl anylose for the stabilization of horseradish peroxidase via addition and cross-linking.* Enzyme and Microbial Technology, 2004. **34**: p. 248-254.
7. Miland, E., M.R. Smyth, and C. O'Fagain, *Increased thermal and solvent tolerance of acetylated horseradish peroxidases.* Biotechnology and Applied Biochemistry, 1996. **19**: p. 63-67.
8. Liu, J.Z., Wang TL, Huang MT, Song HY, Weng LP, Ji LN., *Increased thermal and organic solvent tolerance of modified horseradish peroxidase.* Protein Eng Des Sel, 2006. **19**(4): p. 169-73.
9. O'Brien, A.M., A.T. Smith, and C. O'Fagain, *Effects of phthalic anhydride modification on horseradish peroxidase stability and activity.* Biotechnol Bioeng, 2003. **81**(2): p. 233-40.
10. Garcia, D., F. Ortega, and J. Marty, *Kinetics of thermal inactivation of horseradish peroxidase: stabilizing effect of methoxypoly(ethylene glycol).* Biotechnology and Applied Biochemistry, 1998. **1998**(27): p. 49-54.
11. Bonnaffe, D., M. Therisod, and J. Seris, *Chemical deglycosylation of horseradish peroxidase and surglycosylation using a new glycosylating reagent: effects on catalytic activity and stability,* in *Stability and stabilization of enzymes*, W. van den Tweel, A. Harder, and R. Buitelaar, Editors. 1993, Elsevier: Amsterdam. p. 223-227.
12. Azevedo, A.M., Prazeres DMF, Cabral JMS, Fonseca LP., *Stability of free and immobilized peroxidase in aqueous-organic solvents mixtures.* Journal of Molecular Catalysis B, 2001. **15**: p. 147-153.
13. Fishman, A., Levy I, Cogan U, Shoseyov O., *Stabilization of horseradish peroxidase in aqueous-organic media by immobilization onto cellulose using a cellulose-binding-domain.* Journal of Molecular Catalysis B, 2002. **18**: p. 121-131.
14. Eremin, A.N., Budnikova LP, Sviridov OV, Metelitsa DI. *Stabilization of diluted aqueous solutions of horseradish peroxidase.* Appl Biochem Microbiol 2002; **38**: 151-158.
15. Kuhlmeier, C. and J. Klein, *Stabilisation of enzymes with polyvinylsaccharides I: physical stabilisation of horseradish peroxidase.* Enzyme and Microbial Technology, 2003. **32**: p. 99-106.
16. Morawski, B., S. Quan, and F.H. Arnold, *Functional expression and stabilization of horseradish peroxidase by directed evolution in Saccharomyces cerevisiae.* Biotechnol Bioeng, 2001. **76**(2): p. 99-107.
17. Arnold, F.H. and Z. Lin, *Expression of functional eukaryotic proteins.*, WIPO, Patent application WO 00/006718. 2000.
18. Lipovsek D, Antipov E, Armstrong KA, Olsen MJ, Klivanov AM, Tidor B, Witttrup KD. *Selection of horseradish peroxidase variants with enhanced enantioselectivity by yeast surface display.* Chem Biol. 2007 Oct;14(10):1176-85
19. Steipe, B., Schiller, B., Plückthun, A. and Steinbacher, S. *Sequence statistics reliably predict stabilizing mutations in a protein domain.* J Mol Biol, 1994. **240**(3): p. 188-92.
20. Lehmann, M. and M. Wyss, *Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution.* Curr Opin Biotechnol, 2001. **12**(4): p. 371-5.

21. van den Burg, B. and V.G. Eijnsink, *Selection of mutations for increased protein stability*. *Curr Opin Biotechnol*, 2002. **13**(4): p. 333-7.
22. Lehmann, M., Pasamontes L, Lassen SF, Wyss M, *The consensus concept for thermostability engineering of proteins*. *Biochim Biophys Acta*, 2000. **1543**, 408-415.
23. Ryan, B.J. and R. Barrett, *ProteinParser--a community based tool for the generation of a detailed protein consensus and FASTA output*. *Comput Methods Programs Biomed*, 2007. **85**(1): p. 69-76.
24. Flores, H. and A.D. Ellington, *A modified consensus approach to mutagenesis inverts the cofactor specificity of Bacillus stearothermophilus lactate dehydrogenase*. *Protein Eng Des Sel*, 2005. **18**(8): p. 369-77.
25. Amin, N., Liu AD, Ramer S, Aehle W, Meijer D, Metin M, Wong S, Gualfetti P, Schellenberger V., *Construction of stabilized proteins by combinatorial consensus mutagenesis*. *Protein Eng Des Sel*, 2004. **17**(11): p. 787-93.
26. Miyazaki, J., Nakaya S, Suzuki T, Tamakoshi M, Oshima T, Yamagishi A., *Ancestral residues stabilizing 3-isopropylmalate dehydrogenase of an extreme thermophile: experimental evidence supporting the thermophilic common ancestor hypothesis*. *J Biochem (Tokyo)*, 2001. **129**(5): p. 777-82.
27. Lin, Z., Thorsen T. and Arnold F.H. *Functional expression of horseradish peroxidase in E. coli by directed evolution*. *Biotechnology Progress* 1999, **15** 457-471
28. Grigorenko V, Chubar T, Kapeliuch Y, Borchers T, Spencer F. and Egorov A. (1999). *New approaches for functional expression of recombinant horseradish peroxidase in Escherichia coli*. *Biocatalysis and Biotransformation*, **17**, 359-379.
29. Lei, S.P., Lin HC, Wang SS, Callaway J, Wilcox G., *Characterization of the Erwinia carotovora pelB gene and its product pectate lyase*. *J Bacteriol*, 1987. **169**(9): p. 4379-83.
30. BJ Ryan, B.J. and Ó'Fágáin, C. *Arginine-to-lysine substitutions influence recombinant horseradish peroxidase stability and immobilisation effectiveness*. *BMC Biotechnol*, 2007, **7**: 86.
31. Maniatis, T., E. Fritsch, and I. Sambrook, *Molecular Cloning - a laboratory manual*, ed. C.S.H. Laboratory. 1989, New York: Cold Spring Harbour.
32. Wang, W. and B.A. Malcolm, *Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange Site-Directed Mutagenesis*. *Biotechniques*, 1999. **26**(4): p. 680-682.
33. French, C., E. Keshavarz-Moore, and J. Ward, *Development of a simple method for the recovery of recombinant proteins from the E.coli periplasm*. *Enzyme and Microbial Technology*, 1996. **19**: p. 332-338.
34. Smith A.T, Sanders S.A, Thorneley R.N.F, Burke J.F. and Bray R.C. *Characterisation of a heme active-site mutant of horseradish peroxidase, 41-Phe-Val, with altered reactivity towards hydrogen peroxide and reducing substrates*. *European Journal of Biochemistry*, 1992, **207**, 507-519.
35. Dunford, H., *Heme peroxidases..* 1999, New York: J. Wiley & Sons. 11-12; 92-111.
36. Thompson, J.D., D.G. Higgins, and T.J. Gibson, *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice*. *Nucleic Acids Res*, 1994. **22**(22): p. 4673-80.
37. Duroux, L. and K.G. Welinder, *The peroxidase gene family in plants: a phylogenetic overview*. *J Mol Evol*, 2003. **57**(4): p. 397-407.

38. Yang, Z., S. Kumar, and M. Nei, *A new method of inference of ancestral nucleotide and amino acid sequences*. Genetics, 1995. **141**(4): p. 1641-50.
39. Yang, Z., *PAML: a program package for phylogenetic analysis by maximum likelihood*. Comput Appl Biosci, 1997. **13**(5): p. 555-6.
40. Tams, J.W. and Welinder, K.G. (1998). *Glycosylation and thermodynamic versus kinetic stability of HRP*. FEBS Lett, 1998, **421**: 234-236
41. Gajhede, M., Schuller D, Henriksen A, Smith A, Poulos TL., *Crystal structure of horseradish peroxidase C at 2.15 Å resolution*. Nat Struct Biol, 1997. **4**(12): p. 1032-8.
42. McEldoon J.P. and Dordick J.S. (1996). *Unusual thermal stability of soybean peroxidase*. Biotechnology Progress, **12**, 555-558.
43. Kamal, J.K.A., D.V. Behere. *Activity, stability and conformational flexibility of seed coat soybean peroxidase*. J Inorg Biochem, 2003, **94**, 236-242.
44. Pina, D.G., Shnyrova A.V, Gavilanes F, Rodriguez A, Leal F, Roig M.G, Sakharov I.Y, Zhadan G.G, Villar E. and Shnyrov V.L., *Thermally induced conformational changes in horseradish peroxidase*. Eur J Biochem, 2001. **268**(1): p. 120-6.
45. Ignatenko, O.V., Gazaryan, I.G., Mareeva, E.A., Chubar, T.A., Fechina, V.A., Savitsky, P.A., Rojkova, A.M. and Tishkov, V.I. *Catalytic properties of tryptophanless recombinant horseradish peroxidase*. Biochemistry (Moscow), 2000, **64**, (5), p.583-587.
46. Watanabe, K., Ohkuri, T., Yokobori, S.-I. and Yamagishi, A. *Designing thermostable proteins: ancestral mutants of 3-isopropylmalate dehydrogenase designed by using a phylogenetic tree*. J Mol Biol, 2006. **355**(4): p. 664-74.
47. Eijsink, V.G.H., Bjork A, Gaseidnes S, Sirevag R, Synstad B, van der Berg B, Vriend G., *Rational engineering of enzyme stability*. J Biotechnol, 2004. **113**(1-3): p. 105-20.
48. Magliery, T.J. and L. Regan, *Beyond consensus: statistical free energies reveal hidden interactions in the design of a TPR motif*. J Mol Biol, 2004. **343**(3): p. 731-45.
49. Welinder, K.G., Justesen, A.F.; Kjærsgård, I.V.H.; Jensen, R.B.; Rasmussen, S.K.; Jespersen, H.M.; Duroux, L., *Structural diversity and transcription of class III peroxidases from Arabidopsis thaliana*. Eur J Biochem, 2002. **269**(24): p. 6063-81.
50. Nielsen K-L, Indiani C, Henriksen A, Feis A, Becucci M, Gajhede M, Smulevich, G, Welinder KG., *Differential activity and structure of highly similar peroxidases. Spectroscopic, crystallographic, and enzymatic analyses of lignifying Arabidopsis thaliana peroxidase A2 and horseradish peroxidase A2*. Biochemistry, 2001. **40**(37): p. 11013-21.
51. Lesk, A., *Introduction to Protein Science: architecture, function and genomics*. 2004, New York: Oxford University Press.
52. Kamal, J.K.A, Behere, D.V. *Kinetic stabilities of soybean and horseradish peroxidases*. Biochemical Engineering Journal, 2008, **38**, 110-114.
53. Ryan BJ, Ó'Fágáin C. *Effects of single mutations on the stability of horseradish peroxidase to hydrogen peroxide*. Biochimie 2007, **89** 1029-1032.
54. Carlsson, G.H., Nicholls, P., Svistunenko, D., Berglund, G.I., Hajdu, J. *Complexes of horseradish peroxidase with formate, acetate, and carbon monoxide*. Biochemistry, 2005, **44** 635-642.

## Figure and Table Legends

### **Figure 1 : Phylogenetic tree and ancestral nodes**

(a) Reduced Radial Phylogenetic tree adapted from ref. [37]. A sequence representing each of the 24 clades identified for the peroxidase family. The leaves of the phylogeny represent the following sequences: A1, Aru\_AAA33377; A2, Pba\_CAA66036; A3, Gma\_AAL40127; A4, Pvu\_AAD37430; A5, Nta\_AAA34108; A6, Aru\_CAA40796; F, Ath\_AT4G16270; C, Sol\_AAF63026; B, Ahy\_AAB06183; T, Les\_CAA33852; D, Ghi\_AAL93152; G, Lus\_AAB02926; H, Gma\_AAD11484; M, Tre\_CAA09881; O, Ath\_AtP51; N, Ath\_At4g17960; R, Ath\_At2g24800; P, Ath\_AtP40; I, Ghi\_AAL93154; J, Ath\_AtP35; L, Les\_AAA65637; Q, Ath\_AtP30; K, Ath\_At5g24070; S, Ama\_BAB16317. The letters before the “\_” represent the species names: Aru (*Amoracia rusticana*), Pba (*Populus balsamifera*), Gma (*Glycine max*), Pvu (*Phaseolus vulgaris*), Nta (*Nicotiana tabacum*), Ath (*Arabidopsis thaliana*), Sol (*Spinacia oleracea*), Les (*Lycopersicon esculentum*), Ghi (*Gossypium hirsutum*), Lus (*Linum usitatissimum*), Tre (*Trifolium repens*), Ama (*Avicennia marina*). The number following the “\_” is the Genbank Accession number. “**Archetype**” refers to the earliest class III peroxidase. (b) 3-D structure of extant HRPC A Chain (PDB ID: 1W4Y; ref. 54) *top* and reconstructed extinct archetype peroxidase *bottom*.

**Figure 2:** Alignment of extant HRP (PDB accession code 1W4Y; ref. 54), consensus sequence from approach 1 and reconstructed sequence referred to throughout this paper as the “archetype” (see Figure 1 for phylogenetic positions). Alpha-helices are highlighted by the bars above the alignment and the capital letters refer to the helix classification as found

in ref. [41]. Columns of the alignment selected for mutation in this study are marked by \* beneath the column. Conservation of amino acid positions is noted within the domains across the different sequences.

**Table 1: Stability and Kinetic Characteristics of HRP Consensus Mutants**

Modelled  $k$ -value (first-order exponential decay rate constant), apparent half-life at 50°C ( $t_{1/2}$ ;  $0.693/k$ ) and  $V_m/E$  and apparent  $K'_m$  for each consensus mutant. All stability tests were performed at 0.1 mg protein per mL. ABTS was the reducing substrate for steady-state kinetic analysis.  $k$ -values,  $V_m/E$  and apparent  $K'_m$  values, including standard errors (SE), were calculated using the *Enzfitter*<sup>TM</sup> software package (Version 1.05; Biosoft Ltd, Cambridge, UK: 1987). Data are the mean of three determinations from independent experiments in all cases. “Combined” is the five-fold mutant containing all five amino acid substitutions. Stability analysis was carried out at a constant protein concentration of  $100\mu\text{g.mL}^{-1}$ , whilst kinetic values were determined at a constant 65 pM enzyme per microtitre well.

**Table 1: Stability and Kinetic Characteristics of HRP Consensus Mutants**

Modelled  $k$ -value (first-order exponential decay rate constant), apparent half-life at 50°C ( $t_{1/2}$ ;  $0.693/k$ ) and  $V_m/E$  and apparent  $K'_m$  for each consensus mutant. All stability tests were performed at 0.1 mg protein per mL. ABTS was the reducing substrate for steady-state kinetic analysis.  $k$ -values,  $V_m/E$  and apparent  $K'_m$  values, including standard errors (SE), were calculated using the *Enzfitter*<sup>TM</sup> software package (Version 1.05; Biosoft Ltd, Cambridge, UK: 1987). Data are the mean of three determinations from independent experiments in all cases. “Combined” is the five-fold mutant containing all five amino acid substitutions. Stability analysis was carried out at a constant protein concentration of 100 $\mu$ g.mL<sup>-1</sup>, whilst kinetic values were determined at a constant 65 pM enzyme per microtitre well.

Mutant	Thermal Stability				ABTS Kinetics			
	$k$ -value (min <sup>-1</sup> )	SE (min <sup>-1</sup> )	$t_{1/2}$ (min)	$r^2$ Value	$V_m/E$ (s <sup>-1</sup> )	SE (s <sup>-1</sup> )	$K'_m$ (mM)	SE (mM)
Wild type	0.056	± 0.003	12.4	0.98	482	± 12	0.093	± 0.013
T102A	0.054	± 0.004	12.9	0.97	605	± 37	0.200	± 0.030
Q106R	0.085	± 0.006	8.1	0.96	359	± 7	0.541	± 0.025
Q107D	0.068	± 0.008	10.3	0.96	486	± 33	0.885	± 0.09
T110V	0.051	± 0.005	13.7	0.93	374	± 4	0.170	± 0.02
I180F	0.065	± 0.007	10.7	0.94	497	± 3	0.308	± 0.02
Combined	0.078	± 0.007	8.8	0.92	491	± 10	0.062	± 0.007

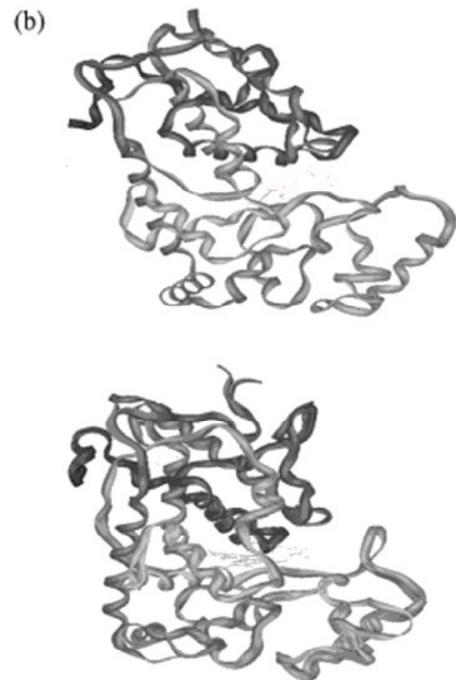
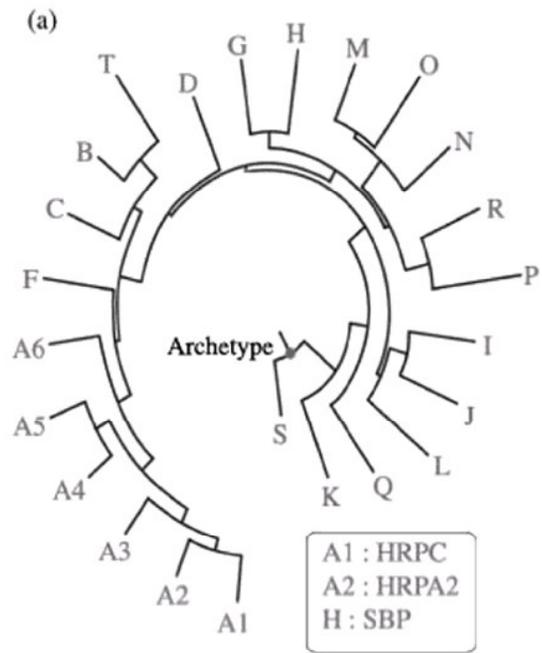


Figure 2

```

HRP -----MQLTPTFYD---NSCPN-----
Approach 1 -----MQL---FY-----CP-----
Archetype MKNLPLFMAMAFSSSSVSLVLVLLLALIASTTTPQAGGYGGDDDDIALASLE DAGLSPNFYKLSKSCPDNGSLLKNVSPDFDLSLESGVA

                                     A           B           C
HRP -----VSNIVRDTIVNELRSDPRIAASILRLHFHDCFVNGCDASILLDNTTSFRT--EKDAFGNANS-ARGF
Approach 1 -----IV-----R--ASLLRLHFHDCFV-GCD-S-LLD-----EK-A-PN-----RGF
Archetype SILTLNKKRNSDKYLRQQLTPEACVBSIVRSVQSAFKSDPTIAAALLRLHFHDCFVNGCDASVLLDPDSDGGNTEBKTAPPNLNLGLRGF

C           D           D'           E           F
HRP PVIDRMKAAVESA--CPRT-VSCADLLTIAAQQSVTLAGGFSWRVPLGRRDSLQAFLLD-ANANLPAPFFTLPLKDSFRNVGLNRSDDLVAL
Approach 1 -VID-IK---E---CP---VSCADILALAARDSVVL-GGF-W-VPLGRRDS--A-----AN--LP-P---L--L---F---GL---DLVAL
Archetype EVIDDIKAALAAAPQCPGVSVCADILALAARDAVVLSSGGPSWEVPLGRRDGLRSAADVANNLPSPTFTVDTLISLFASKGLNNVTDLVAL
                                     *   **   *

F           F'           F''           G           H
HRP SGG-BTFGKNQCRFIMD--RLYNFSNT-GLPDPTLNTTYLQTLRGLCPL-----NGNLSALVDFDLRTPTFDNKYVYNLEBQKGLIQSDQEL
Approach 1 SG--HT-G--C--F--RLYN-----DP-LN-TY-----G-----LD--TP--FDN--YY-NL-----SDQ-L
Archetype SGAGHTIGKAHCSSFSNRFLYNFNNTTGNPDPSLDPTFASQLQALCPQSGNDGNGDDTTVTNLDPGTPNTFDNSYKKNLLSGKLLQSDQEL
                                     *

H           I           J
HRP FSSPN-ATDTIPLVRSFAN-STQTFNAPVEAM-DRMGNITPLTGT--QQQIRLNCRVVNSGS-----
Approach 1 F-----T--V--F-----PF--F--AM--RMGNI--LTG---GEIR--C--VN-----
Archetype FTDPS TKPIVKPFVBSFANGNQLTFEFAEAM-VKMGNI GVL TGKRQGEIRKNCRVVNSRSVATSSVLEEDVEAEVLLAGM

```

Ryan-O’Connell-Ó’Fágáin – Supplementary Material: Comparison of Soybean Peroxidase (SBP) and Horseradish Peroxidase A2 (HRP-A2) Sequences with Archetype

HRP -----MQLTPTFYD---NSCPN-----  
 Approach 1 -----MQL--FY----CP-----  
 Approach 2 -----M-----L-----L--FY----CP-----  
 Archetype MKNLNFNLFMAMAFSSSSVSLLVLVLLLALIASTTTPQAGGYGGDDDDIALASLEDAGLSPNFYKLSKSCPDNGSLLKNVSPDFDSLESQVA

SBP -----QLTPTFYR--ETCPN-----  
 HRP-A2 -----MAVTNLPTCDGLFIISLIVIVSSIFGTS-----SAQLNATFYS--GTCPN-----

HRP -----VSNIVRDTIVNELRSDPRIAASILRLHFHDCFVNGCDASILLDNTTSFRT--EKDAFGNANSA-RGF  
 Approach 1 -----IV-----R--ASLLRLHFHDCFV-GCD-S-LLD-----EK-A-PN----RGF  
 Approach 2 -----E-IVR--V-----D---AA-LLRLHFHDCFV-GCDAS-LLD-----EK-A--N-N---RGF  
 Archetype SILTLNKKRNSDKYLRQQLTPEACVESIVRSVVQSAFKSDPTIAAALLRLHFHDCFVNGCDASVLLDPDSSGGNNTTEKTAPPNLNLGLRGF

SBP -----LFPVIFGVIFDASFTDPRIGASLMRLHFHDCFVQGCDSVLLNNTDTIES--EQDALPNINSI-RGL  
 HRP-A2 -----ASAIVRSTIQALQSDTRIGASLIRLHFHDCFVNGCDASILLDDTGSIQS--EKNAGPNVNSA-RGF

HRP PVIDRMKAAVESA--CPRT-VSCADLLTIAAQSVTLGGPSWRVPLGRRDSLQAFDL-ANANLPAPFFTLQKDSFRNVGLNRSSDLVAL  
 Approach 1 -VID-IK---E---CP---VSCADILALAARDSVVL-GGP-W-VPLGRRDS--A-----AN--LP-P--L--L---F--GL---DLVAL  
 Approach 2 -VID-IK--LE---CP---VSCADILALAARD-V-L-GGP-W-V-LGRRDG-----AN--LP-P-----F--GL---DLVAL  
 Archetype EVIDDIKAALEAAPQCPGVSVSCADILALAARDAVVLSSGGPSWEVPLGRRDGLRSAAADVANNLPSPTFTVDTLISLFASKGLNNVTDLVAL

SBP DVVNDIKTAVENS--CPDT-VSCADILAIAAEIASVLGGPGWPVPLGRRDSLANTL-ANQNLPAFFNLTQLKASFVQGLN-TLDLVT  
 HRP-A2 NVVDNIKTALENA--CPGV-VSCSDVLALASEASVSLAGGWSVTVLLGRRDSLANTLAG-ANSSIPSPIESLSNITFKFSAVGLN-TNDLVAL