Structural analysis of catalase from two *Musa* accessions, FHIA18 and Williams, and from *Ravenala madagascariensis*

Sher-ullah S. S. Khoyratty[1], **Manoel T. Souza Jr.**[2] **and Yasmina Jaufeerally-Fakim**[1]*

[1] Faculty of Agriculture, University of Mauritius, Mauritius

[2] Embrapa LABEX, Europa Plant Research International (PRI), Wageningen University & Research Centre (WUR), Netherlands

* Corresponding author

  Email: yasmina@uom.ac.mu

Abstract

Nucleotide sequences of catalase were obtained following amplification using specific primers and were blasted against *Musa acuminata* catalase 2 mRNA from NCBI (157418810). Clustering of the amino acid sequences from NCBI was done using Clustal X. The latter revealed that FHIA18 catalase is more related to *Ravenala madagascariensis* (*Musa* relative) catalase while the Williams catalase is more related to a clade containing a *Musa acuminata* (*Musa* ancestor) catalase from NCBI. The tertiary structures and the catalase consensus functional sites, based on the *Pseudomonas syringae* catalase structural template, were obtained for FHIA18, Williams, *Ravenala madagascariensis* and *Musa acuminata* catalases. They were found to differ slightly. Using known features of catalase active sites, four pre-requisite criteria were defined to find such sites: (1) Position of tyrosine axial to heme determined by X-ray diffraction, (2) 7 conserved amino acids in the active site found by sequence alignment, (3) favourable docking energy, and (4) presence of an unobstructed long tunnel that leads the ligand to the active site. Two differing potential docking sites were found for both FHIA18 and Williams that fit a maximum number of criteria. In terms of sequence, the region of the docking site for Williams is within the catalase domains as seen upon NCBI blast. The counterpart of FHIA18 for the same region is not. This sequence difference between FHIA18 and Williams affects the best docking site in FHIA18 and Williams *in silico*.

Keywords: catalase, conserved, clustering, active sites, pockets, tunnels, docking

Introduction

Banana is an herbaceous plant of the genus *Musa* and is the developing world's fourth most important food crop (after rice, wheat and maize) (www.traditionaltree.org) [Nelson *et al.*, 2006]. *Musa* variety FHIA18 was bred with parents Prata Enane (AAB) x SH-3142(AA) and has an AAAB genome (Foundation for Agricu Research, Honduras). *Musa* variety Williams is sterile and has an AAA genome [Ploetz *et al.*, 1999].

Ravenala madagascariensis (traveller's palm) is recognised as a relative of *Musa* [Sharrock, 1998]. *Musa acuminata* is one of the ancestors of commercial bananas [Robertson, 2004]. A study was conducted where 35 clones of *Musa acuminata Colla*, representing eight putative subspecies and *M. balbisiana Colla*, were analyzed for isozyme variation of catalase. Subspecies specific alleles have been identified in *Musa acuminata* subspecies *microcarpa*, *burmannica*, *errans*, and *zebrina* [Jarret and Litz, 1986]. Catalase is a major enzyme in the plant defence system against biotic and abiotic stress [Muckenschnabel *et al.*, 2002].

Catalase is a ubiquitous antioxidant enzyme, found in both prokaryotes and eukaryotes, and is involved in the protection of cells from the toxic effects of peroxides. It catalyses the conversion of hydrogen peroxide to water and molecular oxygen [Marchler-Bauer *et al.*, 2007]. Catalase acts in the cell at sites with an electron transport chain and provides protection against oxidative stress caused by hydrogen peroxide. Such sites include the peroxisome [Mittler *et al.*, 2004], the mitochondria and chloroplast. Catalase and ascorbate peroxidase are two enzymes involved in detoxifying hydrogen peroxide in cells. Catalase, however, degrades hydrogen peroxide at an extremely rapid rate.

Multiple catalase isozymes encoded by specific genes are found in plants [Scandalios, 2005], e. g. catalase 1, catalase 2 and catalase 3 occur in *Arabidopsis thaliana* and each catalase member varies in terms of their predominance for different plant organs and in their expression depending on the circumstances and in th number of introns present (lost and gained during evolution) [Frugoli *et al.*, 1998]. A study where catalases are clustered has shown that diverse taxa of bacterial catalases are closely related to the plant catalases. The clustering of catalase amino acid sequences shows that catalase 1, catalase 3 and catalase A from different sources are all found in the same clade [Klotz *et al.*, 1997]. Catalase analogs have also been used to relate to defence markers. Trognitz *et al.*, 2002, studied polymorphic catalase analogs present in resistant potato hybrids and absent in the hybrids susceptible to *Phytophtora infestans*.
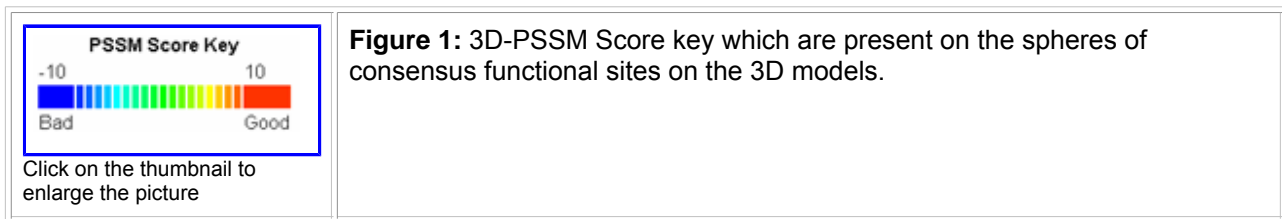
Heme is essential for the enzymatic function of some catalases and is found in the active site of heme-dependent catalases. The degradation of hydrogen peroxide by heme-containing catalases occurs at the heme that lies inwards about 30Å from the protein surface [Amara *et al.*, 2001]. Non-heme catalases originate from prokaryotes [Klotz and Loewen, 2003]. Heme-containing catalases are homotetramers, and each monomer contains a major channel that starts at the protein surface and ends at the deep-buried heme [Amara *et al.*, 2001]. In fact, irrespective of whether they contain heme or not, catalases share a unique structural feature of long channels that provide access to the active sites of these enzymes from the molecular surface. The access channels have crucial roles in the mechanism of enzyme action. Contemporary biochemistry is now becoming increasingly dependent on 3-dimensional structural visualizations to unlock the fundamental process of life [Chelikani *et al.*, 2005].

The heme metal-binding site within the catalase active site has an axial amino acid which is tyrosine (Tyr or 'Y' as one letter code) for several catalase accessions at NCBI (http://www.ncbi.nlm.nih.gov/). The position of the tyrosine amino acid which is axial to heme is conserved or has minor variations when compared to *Bos taurus* (cattle) (NCBI accession: P00432 Residue: 358), *Secale cereale* (rye) catalase (NCBI accession: P55310 Residue: 347), *Triticum aestivum* (bread wheat) (NCBI accession: P55313 Residue: 348), *Zea mays* (NCBI accession: P18122 Residue: 348) and *Oryza sativa* (japonica cultivar-group) (NCBI accession: Q0D9C4 Residue: 348). The conserved domain database (CDD) at NCBI (http://www.ncbi.nlm.nih.gov/) consists of a collection of well-annotated multiple sequence alignment models for ancient domains and full-length proteins. cd00328 catalase sequence cluster from the CDD [Marchler-Bauer *et al.*, 2007] database shows the following conserved residues in the heme binding pocket of all heme catalases: HSNFFRY (His, Ser, Asn, Phe, Phe, Arg, Tyr).

Using the software MODELLER, Sekhar *et al.*, 2006, have built a model of the tertiary structure of a rice catalase using as template the crystal structure of catalase I from *Pseudomonas syringae* (RCSB Protein Data bank (PDB) (http://www.rcsb.org/pdb/home/home.do) [Berman *et al.*, 2000] code: 1m7s).

The consensus functional site is assigned by Phyre on a 3D model based on the PROSITE database of functional motifs. The method is based on statistical estimation of the expected number of a PROSITE pattern to occur in a given sequence [Solovyev and Kolchanov, 1994]. Phyre is the successor of 3D-PSSM but uses the same score key for consensus functional sites that align between the template and the query structure.

The score gives the degree of similarity of the query to the template at any given position depending on the colour. White colour indicates unmatched regions (Fig. 1).

Figure 1: 3D-PSSM Score key which are present on the spheres of consensus functional sites on the 3D models.

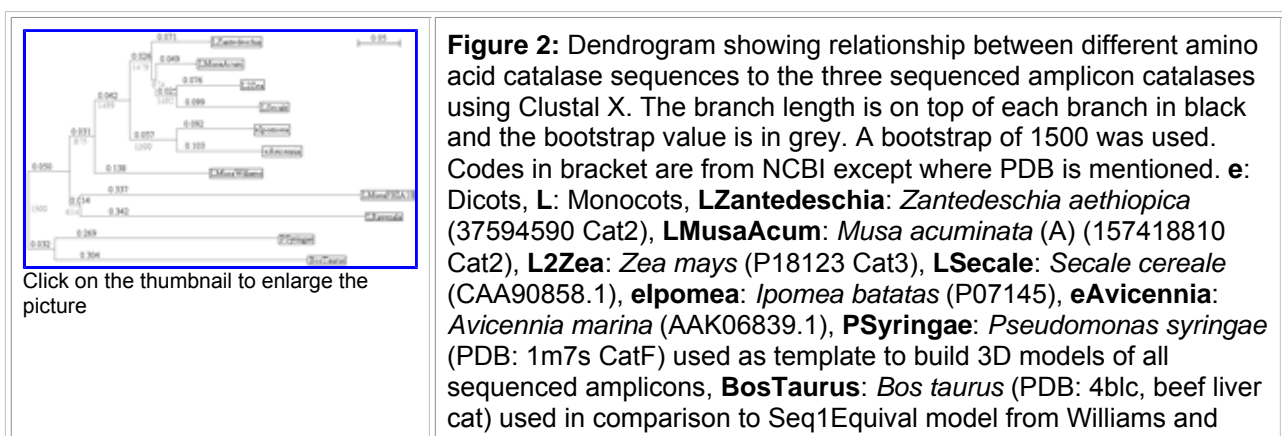Click on the thumbnail to enlarge the picture

Since sterile and economically important banana varieties originate from *Musa acuminata* and *Musa balbisiana*, and given that they are mostly propagated vegetatively, they form an appropriate system that can be used to assess genetic diversity. This study describes the analysis of catalase gene sequences from different varieties in order to have an insight into possible evolutionary polymorphism that could have accumulated. These are compared with an outgroup species related to banana, *Ravenala madagascariensis*.

Methodology

Primer design and PCR: Complete catalase amino acid sequences from NCBI were aligned using MultAlin (http://prodes.toulouse.inra.fr/multalin/multalin.html) [Corpet, 1988] and conserved regions were back-translated with Entelechon (http://www.entelechon.com/backtranslation) using the codon usage table for *Musa acuminata* from Kasuza (http://www.kazusa.or.jp/codon/) [Nakamura *et al.*, 2000]. Forward and reverse primers were selected with Primer3 (http://fokker.wi.mit.edu/primer3/) [Rozen and Skaletsky, 2000] and the primers synthesised by Inqaba Biotechnologies, South Africa. PCR was performed with 1X buffer, 2.0 mM MgCl2, 30 pmol primer, 0.2 mM dNTP, 1.25 units Taq Polymerase and 200 ng DNA in a final volume of 25 μl. All reactants were from Bioline. The cycle used was 95ºC for 5 min, 45 cycles of 95ºC for 30 sec, 53ºC for 30 sec and 72ºC for 2 min. The final cycle was 72ºC for 10 min on a BioRad MyCycler.

Purification and sequencing:
The PCR products were sequenced after purification using Promega Wizard PCR Preps DNA purification system as recommended by the manufacturer. Sequences were assembled using Cap3 (http://pbil.univ-lyon1.fr/cap3.php) [Huang and Madan, 1999]. The sequences were then aligned against a *Musa acuminata* catalase 2 complete mRNA accession 157418809 from NCBI to check for the presence of intron sequences. The sequences were translated and the correct frame found on NCBI.

Clustering catalases: ClustalX ver 1.83 [Thompson *et al.*, 1997] was used with the amplicon sequences and several accessions at NCBI were used to generate a dendrogram. A bootstrap value of 1500 was used (Fig. 2).

Figure 2: Dendrogram showing relationship between different amino acid catalase sequences to the three sequenced amplicon catalases using Clustal X. The branch length is on top of each branch in black and the bootstrap value is in grey. A bootstrap of 1500 was used. Codes in bracket are from NCBI except where PDB is mentioned. **e**: Dicots, **L**: Monocots, **LZantedeschia**: *Zantedeschia aethiopica* (37594590 Cat2), **LMusaAcum**: *Musa acuminata* (A) (157418810 Cat2), **L2Zea**: *Zea mays* (P18123 Cat3), **LSecale**: *Secale cereale* (CAA90858.1), **elpomea**: *Ipomea batatas* (P07145), **eAvicennia**: *Avicennia marina* (AAK06839.1), **PSyringae**: *Pseudomonas syringae* (PDB: 1m7s CatF) used as template to build 3D models of all sequenced amplicons, **BosTaurus**: *Bos taurus* (PDB: 4blc, beef liver cat) used in comparison to Seq1Equival model from Williams and

Click on the thumbnail to enlarge the picture

LMusaWilliams, **LMusaFHIA18** and **LRavenala**: the three sequenced amplicons from Williams (AAA), FHIA18 (AAAB) and *Ravenala madagascariensis* respectively.

Building a 3D model with consensus functional sites: 3D models were built with Phyre (http://www.sbg.bio.ic.ac.uk/phyre/) [Kelley *et al.*, 2000]. The *Pseudomonas syringae* (PDB (http://www.rcsb.org/pdb/home/home.do) [Berman *et al.*, 2000]: 1m7s) template was used, as it had the highest identity and lowest expected value. Sekhar *et al.*, 2006, also used this template to build the 3D model of rice catalase. The results were viewed and edited using RasMol [Sayle and Milner-White, 1995] and PyMol [DeLano, 2002]. Phyre aligns the secondary structures of the *Pseudomonas syringae* (PDB: 1m7s) template to the subject catalase and identical structures known to be associated with function from PROSITE (http://www.expasy.ch/prosite/) [Hofmann *et al.*, 1999] are assigned as consensus functional sites by Phyre. The consensus sites were included in the tertiary structure by using the consensus functional site script of RasMol.

Finding active site (HSNFFRY) in a heme-binding pocket: Amino acids in FHIA18 and Williams catalase that aligned to the known conserved amino acids of the heme-binding pockets in the sequences of the catalases from *Zantedeschia aethiopica* (NCBI accession: AAG61140.2) and *Bos taurus* (cattle) (NCBI accession: P00432) were found using MultAlin. Using the tyrosine known to be axial to heme, determined alignment, the other 6 conserved amino acids were sought in the close vicinity of this tyrosine for FHIA18 and Williams using PyMol in slab mode.

Docking   ligand: The hydrogen peroxide ligand 3D model was obtained from RCSB Protein Data bank (PDB) [Berman *et al.*, 2000]. Using Quantum (http://q-pharm.com/) [Fedichev *et al.*, 2005], hydrogen peroxide ligands were docked in 30Å X 30Å X 30Å (x, y, z) boxes covering the FHIA18 sequence related to catalase (FHIA18 Seq1) and a 20Å X 22Å X 25Å (x, y, z) box covering a region that encompass the central residue 314 (tyrosine) in Williams. Ten positions in Williams around TYR-314 and 120 positions in FHIA18 were checked for the energetically favourable docking site. The most energetically favourable one that also had the 7 catalase active site conserved amino acids in close vicinity, was chosen.

Finding pockets and building tunnels: Using CASTp (http://sts-fw.bioengr.uic.edu/castp/calculation.php) [Binkowski *et al.*, 2003] with spheres of 1.4Å in diameter, 39 pockets in Williams and 24 pockets in FHIA18 with at least one access to the outside mouth were found. Using PyMol [DeLano, 2002] with a CASTp plugin [Binkowski *et al.*, 2003], the pockets were mapped on to the Williams and FHIA18 catalase with the 'surface option' on PyMol [DeLano, 2002]. The docked position was found in the model to check whether a tunnel leads from such a position to the outside. For Williams, PyMol was used in slab option to locate the tunnel.

Structural   pairwise   alignment: The models of FHIA18 and Williams were superimposed with MATRAS 3D pairwise structural alignment (http://biunit.aist-nara.ac.jp/matras/matras_pair.html) [Kawabata, 2003].

Finding polymorphic regions related to catalase domains: Alignments were performed with FHIA18 and Williams catalase using MultAlin [Corpet, 1988] for a region encompassing the *in silico* docking site of FHIA18 found. The region was protein-protein blasted at NCBI 1998, for both FHIA18 and the Williams equivalent. The sequence of the region was then converted into 3D models using Phyre for comparative purposes.


Results

Using catalase-specific primers ATC CTG CTG GAG GAC TAC CA (forward) and GTT CAC CTC CTC GTC GCT GT (reverse) PCR products of size 1064 bp, 1053 bp and 1054 bp were obtained from FHIA18, Williams and *Ravenala madagascariensis* respectively. The sequences of FHIA18, Williams and *Ravenala madagascariensis* amplicons blast with many catalases with low E-value and high % identity at NCBI with the first hit being a *Musa acuminata* NCBI: 157418810 accession with 98% id, E = 0.0 for FHIA18; 97% id, E =

0.0 for Williams and 88% id, E = 0.0 for *Ravenala madagascariensis*. The first hit is a complete mRNA sequence. The first hit was aligned to the three sequenced amplicons in MultAlin to check for the presence introns. There were none.

Building and comparing 3D models of the amplicons

A 3D model was built for the three catalase sequences FHIA18, Williams, *Ravenala madagascariensis* and the NCBI accession 157418810 *Musa acuminata* with Phyre using the same template *Pseudomonas syringae* SCOP: d1m7sa_ model to allow for comparison with E-values 2.9e-37, 8e-36, 4.3e-40 and 0 respectively, % i.d: 29%, 41%, 28% and 44% respectively and an estimated precision of 100% in all cases. The results at Phyre show that all three amplicon sequence models blast well with catalases with low E-values, high % i.d and a high estimated precision including *Bos taurus* SCOP: d4blca_. 3D positions of consensus functional sites, relative to the same template, differ on the FHIA18, Williams, *Ravenala madagascariensis* and *Musa acuminata* models (Fig. 3, Plates A-D). The 3D model of *Musa acuminata* (Fig. 3, Plate D) differs from FHIA18, Williams and *Ravenala* (Fig. 3, Plates A, B and C) despite having a high percentage identity to the three amplicons upon NCBI blast. A structural alignment between *Ravenala* and *Musa acuminata* on MATRAS 3D pairwise alignment shows the *Ravenala* model (Fig. 3, Plate C) is similar to a structure within the *Musa acuminata* model (Fig. 3, Plate D).
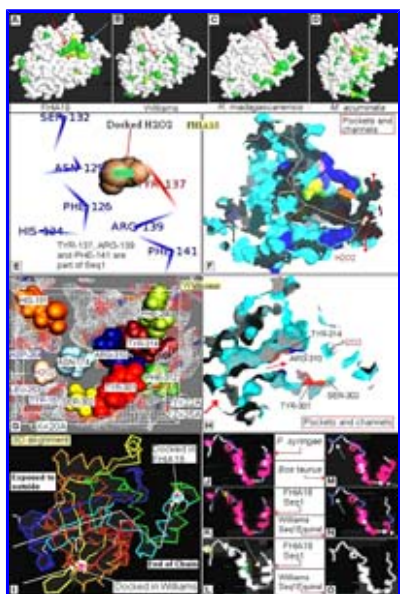
Click on the thumbnail to enlarge the picture

Figure 3: Plates A-D: The consensus functional sites relative to *Pseudomonas syringae* d1m7sa_ are different between *Musa* (FHIA18), *Musa* (Williams), the *Musa* relative *Ravenala madagascariensis* and the ancestor *Musa acuminata* NCBI 157418810 (lemon green, green, dark green and orange sphere based on the 3D-PSSM score key). The regions where most consensus functional sites cluster in 3D are shown with a red arrow. The blue arrow in Plate A shows where hydrogen peroxide was docked in FHIA18 which is on the boundary of the consensus functional site cluster. Plate E: A ligand docking site in FHIA18 which satisfies the maximum number of criteria for the active site. The ligand docks on TYR-137 with Quantum. The site contains the 7 conserved amino acids in a catalase active site in close vicinity. Three of the amino acids (Tyr137, Arg139 and Phe141) are found on a structure which will be named 'Seq1' and will be further studied. Plate F: The docking site in plate E in relation to the whole enzyme showing an unobstructed tunnel leading from the outside to the docking site (white lines). The tunnel is closed to the outside but open on the inside when checked on Quantum with respect to the docking site. Red arrows show closed routes further inside the enzyme (checked on Quantum and PyMol). The blue surface structures are amino acids involved in making pockets from CASTp while all colours other than marine blue are amino acids involved in the consensus functional sites plate G: A ligand docking site in Williams which satisfies the maximum number of criteria for the active site. The ligand docks on Tyr187 with Quantum which is close to Tyr314, a conserved tyrosine associated with catalase active sites. The site contains the 7 conserved amino acids in a catalase active site in close vicinity (a distance of not more than 20Å). Plate H: The docking site in Plate G in relation to the whole enzyme showing an unobstructed tunnel leading from the outside to the docking site (red arrows). Plate I: Structural pairwise alignment of the main chains of FHIA18 and Williams catalase. Only the structures that align have been shown. Structures that do not align have been masked. The white spheres with a purple dot on the superimposed model shows the amino acids on which the ligands have docked in FHIA18 (Tyr137) and Williams (Tyr187). The white lines show the route of the ligand to the active site. d1m7s_ template is composed of 4 similar chains. The 4 chains of d1m7sa_ meet at the tip close the docking site in FHIA18 which is the end of the chain. Plate J-O: Magenta: α-helices, pale blue: turns, white: other residues, and green:

consensus functional sites. Seq1 model on which the ligand has docked in FHIA18 (residue 135 to 171) and equivalent structures from Williams (residue 135 to 168), *Pseudomonas syrinagae* PDB code: d1m7sa_ (used as template to build all the catalases) 46% identity to Seq1 and *Bos taurus* PDB code: d4blca_ 49% identity to Seq1. Seq1 blasts with low *E*-values and high identities to catalases at NCBI whereas the same counterpart in Williams does not blast to any gene at NCBI. Despite this, the structure of this polymorphic region in FHIA18 (Seq1) resemble that of the Williams counterpart (Seq1Equival) (Plates K-O) but Seq1 has three consensus functional sites Tyr167, Phe154 and His146 (green plate L) while Seq1Equival has no such sites. The *Bos taurus* (plate M) structure blasts to catalases while resembling more Williams than FHIA18. On the other hand, the *Pseudomonas syringae* (plate J) model resembles more FHIA18 than Williams.

In all cases the models were built with *Pseudomonas syringae* SCOP: d1m7sa_ as the chosen template at Phyre with one of the highest identity, highest % precision and lowest *E*-value. The template is defined by a fold/PDB descriptor, superfamily and family of heme-dependant catalase-like proteins and its structure determined by X-ray diffraction.

Finding the docking sites in FHIA18 and Williams catalase

Four criteria were used to find the active site in the Williams and FHIA18 catalases. The region that fits maximum number of criteria was chosen as the active site. The four criteria were: (1) Axial tyrosine to heme is present at conserved position compared to catalases for which the position has been determined by diffraction, (2) 7 active site conserved amino acids are present in the vicinity [Sekhar *et al.*, 2006], (3) the energy for the ligand to dock at the docking site is favourable [Sekhar *et al.*, 2006], and (4) an unobstructed long passage that can lead the ligand from the outside to the active site, which is deep-seated [Amara *et al.*, 2001; Chelikani *et al.*, 2005].

Sequence polymorphism in the docking site conserved amino acids of *Bos taurus* and *Zantedeschia aethiopica* catalases relative to FHIA18 and Williams

FHIA18 and Williams sequences were aligned to two catalase sequences (*Bos taurus* and *Zantedeschia aethiopica*) for which the position of the 7 conserved amino acids in the catalase active site are known from NCBI.

The results in Fig. 4 show that for FHIA18 three out of 7 and for Williams five out of 7 conserved positions are the same as *Zantedeschia aethiopica* and *Bos taurus* catalases. Since the 7 amino acids should be present in the vicinity of an active site and that such residues are generally conserved across organisms, using the coordinates of the tyrosine (Y) at position 360 in Fig. 4, the 6 other amino acids were sought in 3D, using PyMol with slab mode, for FHIA18 and Williams. Position Tyr360 in the alignment in Fig. 4 is amino acid Tyr314 in the linear sequence of Williams. All 7 amino acids were located around tyrosine 314 in the case of Williams (Fig. 3, Plate G) but not in the case of FHIA18 and this despite checking in 3D around all other aligned positions other than 360 (Fig. 4) in the alignment that corresponds to the aligned conserved amino acids to *Zantedeschia aethiopica* and *Bos taurus.*

Click on the thumbnail to enlarge the picture

Figure 4: Amino acid sequence alignment with MultAlin to find the relative positions of amino acids in Williams and FHIA18 catalases to those known to be present in the active sites of *Zantedeschia aethiopica* (NCBI accession: AAG61140.2) and *Bos taurus* (cattle) (NCBI accession: P00432) catalases. Marked positions in green and violet are amino acids involved in the heme-binding pocket of *Zantedeschia aethiopica* and *Bos taurus* obtained from NCBI and calculated at the aligned positions 75, 114, 148, 153, 161, 356 and 360.

Docking energy and pockets within FHIA18 and Williams catalase

Apart from the presence of the 7 conserved amino acids in a catalase active site and the presence of the aligned tyrosine axial to heme, two other criteria were chosen to further test the docking site of Williams. The site with the most favourable docking energy was sought using Quantum (Tab. 1 and Fig. 3, Plates G and H) and the presence of pockets and open tunnels leading from the outside to the active site were calculated using CASTp and implemented on the model with PyMol in surface mode (Fig. 3, Plate H).

Table 1: Hydrogen peroxide ligand docking energy for FHIA18 and Williams.

Energy Measurements	FHIA18	Williams
IC_{50}, µmol/l	0.135	0.135
E_{bind}, kJ/mol	−7.85	−8.09
E_{es}, kJ/mol	−14.97	−15.38
E_{tor}, kJ/mol	4.93	−0.32

In both cases, the IC_{50} value shows that the docking energy is favourable for the site to dock the ligand. IC_{50}, µmol/l - inhibition constant (IC) value; E_{bind}, kJ/mol - free binding energy which is equal to the sum of all listed contributions (E_{es}, E_{vdw}, TdS and E_{tor}). The lower the value, the better the protein-ligand interaction; E_{es}, kJ/mol - electrostatic and solvation energy and E_{tor}, kJ/mol - ligand internal energy change.

Similarly, the docking site which satisfied most of the criteria for a catalase active site was sought in the case of FHIA18 with a different method. Twelve different regions were assessed for docking with Quantum with each region having 10 docking trials i. e. 12 × 10 = 120 potential docking sites. The best docking site in each of the twelve regions was assessed for the presence of all 7 conserved amino acids. In only one case out of the twelve, all 7 amino acids were found in close vicinity (Fig. 3, Plate E) while the ligand docked onto tyrosine 137 (tyrosine is an axial amino acid to heme in an active catalase site).

Superimposition of FHIA18 and Williams catalase relative to the docking sites

To assess the difference in structures between FHIA18 and Williams in relation to the docking sites of the ligands, the models of FHIA18 and Williams were structurally aligned with MATRAS 3D pairwise alignment and the docking sites were located on the superimposed model (Fig. 3, Plate I). The superimposed model is composed of most of FHIA18 and Williams structures i. e. overall, a high level of structural similarity exist between both.

Sequence polymorphism in relation to the docking sites of FHIA18 and Williams

The docking site in FHIA18 was further investigated by aligning the 1D sequences of Williams and FHIA18 whereby several polymorphic regions between the two amplicons were found. One such region blasts to catalases at NCBI with low E-value and high identity for FHIA18 (Seq1: QEY WRx FDF xSH HPx SLx TFF

FxF DDV GVP SDY RxM E) but not for the equivalent sequence in Williams (Seq1Equival: GVL EGx RLP xAP PRx PPx LLL PxR RRG RPV RLP P) nor is the Williams sequence related to any sequence at NCBI. This sequence region was converted into its corresponding model with Phyre and it was found that the ligar docks on the model in FHIA18 (Fig. 3, Plate L). Similarly, a second amino acid sequence was found which related to catalase domains upon NCBI blast and Phyre search but this time in the case of Williams (Seq2: RVF AYG DTQ) but not for the equivalent in the case of FHIA18 (Seq2Equival: QGV RVW RHA).

Click on the thumbnail to enlarge the picture

Figure 5: Sequence alignment with MultAlin showing a difference between Williams and FHIA18 catalase (region in red 204-240) in a region close to the position where the ligand has docked in FHIA18 and region 382-390. For region 204-240 FHIA18 Seq1 blasts with catalases whereas Williams Seq1 Equival does not whereas for region 382-390 Williams Seq2 blasts with catalases whereas FHIA18 Seq2 Equival does not. Tyr137 is the amino acid where hydrogen peroxide has docked on FHIA18 (Fig. 3, Plates E, F and I). Tyr314 is an amino acid close to which hydrogen peroxide has docked on Williams (Fig. 3, Plates G, H and I). It is also the aligned position of a conserved tyrosine (Y) in catalases which is associated with heme deduced from X-ray diffraction for the *Bos* aligned accession. FHIA18 has no tyrosine (Y) at this position and has a valine (V) instead. The catalase abbreviations are (accessions codes at NCBI are shown): FHIA18 is FHIA18 catalase, Williams is Williams catalase, Oryza1 is *Oryza sativa* (indica cultivar-group) CAA43814.1, Secale is *Secale cereale* (rye) P55310, Zantedeschia is *Zantedeschia aethiopica* AAG61140.2, Triticum is *Triticum aestivum* (bread wheat) P55313, Zea is *Zea mays* P18122, Oryza2 is *Oryza sativa* (japonica cultivar-group) Q0D9C4 and Bos is *Bos taurus* (cattle) P00432. Although the majority of catalases of plant origin show conserved sequences in the Seq1 region, Williams does not. However, in between the non-conserved amino acid sequences in Williams, there are a few conserved sequences that are also present compared to other catalases from plant origin.

Structure of docking site polymorphic sequences in FHIA18 and Williams compared to equivale structures in other catalases

The models of FHIA18 Seq1 and Williams Seq1Equival have some similarity (Fig. 3, Plates K and N) despite a major difference in sequence (Fig. 5). However the difference between Seq1 and Seq1Equival is that S contains three consensus functional sites whereas Seq1Equival has none (Fig. 3, Plates L and O). WilliamsSeq1Equival has 3 α-helices whereas FHIA18Seq1 has 4 α-helices. The *Bos taurus* model in Fig. 3 ,Plate M, has 49% structural identity to FHA18 Seq1 (Fig. 3. Plate K). The *Pseudomonas syringae* model in Fig. 3, Plate J, has 46% identity to FHIA18 Seq1 from the structural similarity result at Phyre. In contrast, th partial Bos taurus model related to catalase (Fig. 3, Plate M) is more structurally similar to Williams Seq1Equival (Fig. 3, Plate N) especially at two structures (shown with arrow Fig. 3, Plates M and N) and this despite FHIA18 Seq1 being structurally related to catalases, as determined by Phyre, but not Willian Seq1Equival.

Williams Seq2 structurally resembles FHIA18 Seq2Equival. Both models are packed with consensus functional sites that however differ in position when Seq2 and Seq2Equival are compared (Fig. 6, Plate A and B).

Musa acuminata catalase has a similar sequence and structure to the FHIA18 catalase docking region

Although the best docking site for Williams and FHIA18 catalase are not the same (Fig. 3, Plates E, F, G and H), the amino acid sequence alignment (Fig. 5) shows that the sequence of FHIA18 Seq1 is not unique to FHIA18 and occurs in several other accessions including in *Musa acuminata* (alignment result not shown). As a consequence, the 7 conserved amino acids in the heme binding site were also checked in the structu

equivalent to FHIA18 Seq1 in *Musa acuminata* built on Phyre and the position of which was determined by MultAlin 1D sequence alignment. All 7 conserved amino acids were found in close vicinity to tyrosine 169 which is equivalent to tyrosine 137 in FHIA18 (Fig. 6, Plate C and D).
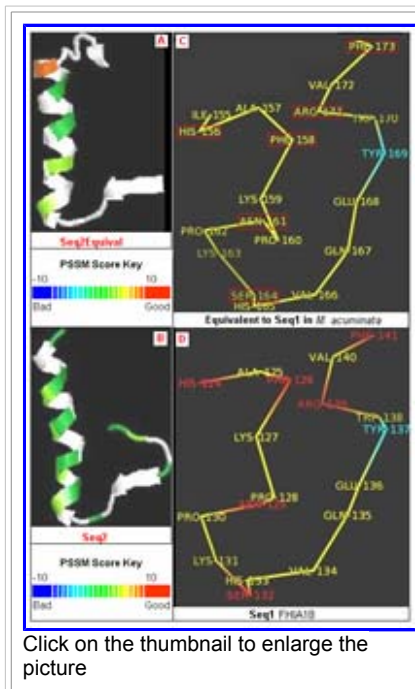

Figure 6: Models of Seq2 (Plate B) in Williams, Seq2Equival in FHIA18 (Plate A) with consensus functional sites compared to d1m7sa_ (green, lemon green and orange). PSSM Score Key indicates degree of similarity between query sequence of FHIA 18 to that of *Pseudomonas syringae* d1m7sa_ catalase template. Seq1 in FHIA18 (Plate D) and the corresponding structure to Seq1 in *Musa acuminata* (Plate C) both have the 7 conserved amino acids in a heme-binding pocket.

Click on the thumbnail to enlarge the picture

Discussion

Musa acuminata is an ancestor to FHIA18 and Williams varieties from which the catalase was probably inherited while *Ravenala madagascariensis* is a relative and shares a more ancient ancestry to the three. Williams and FHIA18 catalase sequences share a 97% and 98% identity to *Musa acuminata* catalase respectively while *Ravenala madagascariensis* share only 88% identity to the same *Musa acuminata* catalase. The Clustal X dendrogram of the amino acids sequences (Fig. 2), shows that Williams catalase clusters in a clade containing the *Musa acuminata* catalase accession while the FHIA18 catalase is in a different clade with *Ravenala madagascariensis* catalase sequence.

The best structural hits obtained at Phyre for the amplicons of Williams, FHIA18 and *Ravenala madagascariensis* sequenced, were a *Pseudomonas syringae* (PDB: 1m7s CatF) and a Bos taurus (PDB: 4blc, beef liver cat) catalase model. The model of *Pseudomonas syringae* (PDB: 1m7s CatF), was used as template to build the Williams, FHIA18, *Ravenala madagascariensis* and *Musa acuminata* models. Sekhar *et al.*, 2006, used the same template to build the model of catalase from rice.

The best hits of nucleotide and amino acid of NCBI blast and the Phyre structural hits are all catalases with low e-values and high % identity. From the dendrogram (Fig. 2), it can be seen that the catalases do not always cluster on the basis of monocotyledonous or dicotyledonous origin. *Ipomea batatas* and *Avicennia marina* catalases both originate from dicotyledons and cluster with catalases from monocotyledons, including those from the *Musa* sequences.

Despite a difference in consensus functional sites, a high level of structural similarity was found between FHIA18 and Williams catalase (Fig. 3, Plate I). The aligned superimposed model generated using MATRA showed that most of the structures present on both FHIA18 and Williams aligned well. The models of the FHIA18, Williams and *Ravenala madagascariensis* amplicons are similar to *Musa acuminata* (Fig. 3, Plate A-D) although some differences are evident. This is because the amino acid sequence of *Musa acuminata* is

complete as it came from a cDNA clone while the other three sequences are nearly complete. The primers used for their amplification span nearly the whole catalase sequence. Structural alignment of Ravenala and *Musa acuminata* sequences on MATRAS 3D pairwise alignment shows that the *Ravenala* model (Fig. 3, Plate C) is similar to part of the *Musa acuminata* model (Fig. 3, Plate D).

The docking site in FHIA18 encompasses Seq1 region (Fig. 3, Plate L, yellow arrow). FHIA18 Seq1 is related to catalase domains (NCBI blast) whereas its counterpart in Williams (Seq1Equival) is not. Despite this, Williams Seq1Equival is similar in structure to FHIA18 Seq1 and parts that differ are present in other catalases such as *Pseudomonas syringae* and *Bos taurus* (Fig. 3, Plates J, K, M and N). This suggests that Williams Seq1Equival could be a mutant of a conserved catalase functional site that has probably lost its function (lacks the three consensus catalase functional sites present on FHIA18Seq1) (Fig. 3, Plates L and O).

Four criteria were used to define the docking site and the site that satisfies most of them was chosen. The *Pseudomonas syringae* d1m7sa_ template used in building the 3D models has a high level of identity to the *Musa* and *Ravenala* *madagascariensis* sequences. The template is defined as a heme-dependant catalase at PDB. For this reason, it was assumed that the catalase sequences in this study are heme-containing catalases. Under this assumption, the first criterion was the presence, obtained by sequence alignment, of a tyrosine relative to a tyrosine known to be axial to heme as determined by X-ray diffraction. The second criterion was the presence of all the other conserved amino acids in close vicinity to the tyrosine. The third criterion was the ability for hydrogen peroxide to dock in such a region with a favourable energy as expected in an enzyme active site. Finally, the fact that active sites of catalases are deep seated and accessible through long channels, which are thought to determine the rate of catalysis [Amara *et al.*, 2001; Chelikani *et al.*, 2005] was an important consideration. Showing that this is the case would further confirm the authenticity of the docking site.

Based on the four criteria, a site was found in Williams where the tyrosine aligned to the tyrosine determined by X-ray diffraction as being axial to heme in the accession *Bos taurus* (Figs. 4 and 5) and it has the other 6 conserved amino acids in its vicinity (Fig. 3, Plate G). A ligand docked in the region with favourable docking energy (Tab. 1) and an open channel (Fig. 3, Plate H) was found that lead the ligand to the docking site. However, the site does not satisfy the last criterion entirely. The active site is not deep seated nor does the ligand have to travel a long channel to reach it (Fig. 3, Plates H and I). In FHIA18, the alignment to the *Bos taurus* accession shows no tyrosine and for each conserved position to several catalases (Figs. 4 and 5), no complete set of 7 conserved amino acids was found in close vicinity in 3D. For FHIA18, searching and docking revealed a site where all but one criteria were met i. e. favourable docking energy (Tab. 1), presence of the 7 amino acids in close vicinity and a long unobstructed channel leading from the outside of the enzyme to the inside (Fig. 3, Plates E, F and I). However, the tyrosine on which the ligand docked (Fig. 3, Plate E) did not align to the tyrosine axial to heme in *Bos taurus* (Fig. 5). Both sequence regions around the docking sites of FHIA18 and Williams are highly related to catalase domains upon NCBI blast. However, where the Williams sequence in the docking region is related to catalases (Seq2 Fig. 5), FHIA18 is not (Seq2Equival Fig. 5) and where the FHIA18 sequence in the docking region is related to catalases (Seq1 Fig. 5), Williams is not (Seq1Equival Fig. 5).

Other accessions used in the alignment show that *Oryza*1, *Secale*, *Zantedeschia*, *Triticum*, *Zea*, *Oryza*2 and *Bos* catalases all have both the sequence of Williams and that of FHIA18 which are related to catalase domains (Fig. 5).

The *in silico* docking site for FHI18 has a long unobstructed channel leading to the active site. For Williams, the ligand still has to go through a long unobstructed channel, but the active site is not deep seated. The latter is similar to that of *Bos taurus*.

The catalases from two banana varieties whose amino acid sequences blast best with the same accession have potentially two different *in silico* docking sites. This could be evidence of the evolution of the catalase gene within the *Musa* species, as a

result of different mutations and/or selection pressures. Comparative studies of other oxidative stress-response genes within FHIA18 and Williams as well as from other *Musa* accessions would give an insight into the dynamic genetic changes that could have occurred in such vegetatively propagated crops.

References

- Amara, P., Andreoletti, P., Jouve, H. M. and Field, M. J. (2001). Ligand diffusion in the catalase from *Proteus mirabilis*: A molecular dynamics study. Protein Sci. **10**, 1927-1935.

- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000). The Protein Data Bank. Nucleic Acids Res. **28**, 235-242.

- Binkowski, T. A., Naghibzadah, S. and Liang, J. (2003). CASTp: Computed Atlas of Surface Topography of proteins. Nucleic Acids Res. **31**, 3352-3355.

- Chelikani, P., Ramana, T. and Radhakrishnan, T. M. (2005). Catalase: A repertoire of unusual features. Indian J. Clin. Biochem. **20**, 131-135.

- Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res. **16**, 10881-10890.

- DeLano, W. L. (2002). The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA.

- Fedichev, P. O., Men'shikov, L. I. and Vinnik, A. A. (2005). Novel ab initio molecular modeling tools for drug discovery. Quantum Pharmaceuticals Ltd., Usievicha str, Moscow, Russia. http://www.q-pharm.com/home/contents/sci_and_tech/science.

- Frugoli, J. A., McPeek, M. A., Thomas, T. L. and McClung, C. R. (1998). Intron loss and gain during evolution of the catalase gene family in angiosperms. Genetics **149**, 355-365.

- Hofmann, K., Bucher, P., Falquet, L. and Bairoch, A. (1999). The PROSITE database, its status in 1999. Nucleic Acids Res. **27**, 215-219.

- Huang, X. and Madan, A. (1999). CAP3: A DNA sequence assembly program. Genome Res. **9**, 868-877.

- Jarret, R. L. and Litz, R. E. (1986). Enzyme polymorphism in *Musa acuminata Colla*. J. Hered. **77**, 183-188.

- Kawabata T. (2003). MATRAS: a program for protein 3D structure comparison. Nucleic Acids Res. **31**, 3367-3369.

- Kelley, L. A., MacCallum, R. M. and Sternberg M. J. E. (2000). Enhanced genome annotation using structural profiles in the program 3D-PSSM. J. Mol. Biol. **299**, 499-520.

- Klotz, M. G. and Loewen, P. C. (2003). The molecular evolution of catalatic hydroperoxidases: evidence for multiple lateral transfer of genes between prokaryota and from bacteria into eukaryota. Mol. Biol. Evol. **20**, 1098-1112.

- Klotz, M. G., Klassen, G. R. and Loewen, P. C. (1997). Phylogenetic relationships among prokaryotic and eukaryotic catalases. Mol. Biol. Evol. **14**, 951-958.

- Marchler-Bauer, A., Anderson, J. B., Derbyshire, M. K., DeWeese-Scott, C., Gonzales, N. R., Gwadz, M., Hao, L., He, S., Hurwitz, D. I., Jackson, J. D., Ke, Z., Krylov, D., Lanczycki, C. J., Liebert, C. A., Liu, C., Lu, F., Lu, S., Marchler, G. H., Mullokandov, M., Song, J. S., Thanki, N., Yamashita, R. A., Yin, J. J., Zhang, D. and Bryant, S. H. (2007). CDD: a conserved domain database for interactive domain family analysis. Nucleic Acids Res. **35**, D237-240.

- Markus, F. (2006). Backtranslation Tool. Entelechon GmbH, Regensburg, Germany. http://www.entelechon.com/bioinformatics/btcited.php?lang=eng.

- Mittler, R., Vanderauwera, S., Gollery, M. and Van Breusegem, F. (2004). Reactive oxygen gene network of plants. Trends Plant Sci. **9**, 490-498.

- Muckenschnabel, I., Goodman, B. A., Williamson, B., Lyon, G. D. and Deighton, N. (2002). Infection of leaves of *Arabidopsis thaliana* by *Botrytis cinerea*: changes in ascorbic acid, free radicals and lipid peroxidation products. J. Exp. Bot. **53**, 207-214.

- Nakamura, Y., Gojobori, T. and Ikemura, T. (2000). Codon usage tabulated from the international DNA sequence databases: status for the year 2000. Nucleic Acids Res. **28**, 292.

- Nelson, S. C., Ploetz, R. C. and Kepler, A. K. (2006). Musa species (banana and plantain). Species Profiles for Pacific Island Agroforestry, www.traditionaltree.org.

- Ploetz, R. C., Haynes, J. L. and Vásquez, A. (1999). Evaluation of bananas for niche markets in subtropical Florida. InfoMusa **8 (2)**, 15-18.

- Robertson, H. (2004). Musa. Iziko Museums of Cape Town, South Africa.

- Rozen, S. and Skaletsky, H. J. (2000). Primer3 on the WWW for general users & for biologist programmers. *In:* Bioinformatics Methods & Protocols: Methods in Molecular Biology, Krawetz, S. and Misener, S. (eds.), Humana Press, Totowa, NJ, pp. 365-386.

- Sayle, R. A. and Milner-White, E. J. (1995). RASMOL: biomolecular graphics for all. Trends Biochem. Sci. **20**, 374.

- Scandalios, J. G. (2005). Oxidative stress: molecular perception and transduction of signals triggering antioxidant gene defenses. Braz. J. Med. Biol. Res. **38**, 995-1014.

- Sekhar, P. N., Kishor, P. B. K., Reddy, L. A., Mondal, P., Dash, A. K., Kar, M., Mohanty, S. and Sabat, S. C. (2006). *In silico* modeling and hydrogen peroxide binding study of rice catalase. In Silico Biol. **6**, 0041.

- Sharrock, S. (1998). The banana and its relatives. *In:* Networking Banana and Plantain: INIBAP Annual Report 1997. International Network for the Improvement of Banana and Plantain, Montpellier, France, pp. 52-55.

- Solovyev, V. V. and Kolchanov, N. A. (1994). Search for functional sites using consensus. *In*: Computer analysis of genetic macromolecules, Kolchanov, N. A. and Lim, H. A. (eds.), World Scientific, pp. 16-21.

- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G. (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25**, 4876-4882.

- Trognitz, F., Manosalva, P., Gysin, R., Niño-Liu, D., Simon, R., Herrera, M. D. R., Trognitz, B., Ghislain, M. and Nelson, R. (2002). Plant Defense Genes Associated with Quantitative Resistance to Potato Late Blight in *Solanum phureja* x Dihaploid *S. tuberosum Hybrids*. Mol. Plant Microbe Interact. **15**, 587-597.