



Article

ROS Consumers or Producers? Interpreting Transcriptomic Data by AlphaFold Modeling Provides Insights into Class III Peroxidase Functions in Response to Biotic and Abiotic Stresses

James New¹, Daniel Barsky² and Claudia Uhde-Stone^{1,*}

¹ Department of Biological Sciences, California State University, East Bay, Hayward, CA 94542, USA; jnew@horizon.csueastbay.edu

² Department of Physics, California State University, East Bay, Hayward, CA 94542, USA; daniel.barsky@csueastbay.edu

* Correspondence: claudia.stone@csueastbay.edu

Abstract: Participating in both biotic and abiotic stress responses, plant-specific class III peroxidases (PERs) show promise as candidates for crop improvement. The multigenic PER family is known to take part in diverse functions, such as lignin formation and defense against pathogens. Traditionally linked to hydrogen peroxide (H₂O₂) consumption, PERs can also produce reactive oxygen species (ROS), essential in tissue development, pathogen defense and stress signaling. The amino acid sequences of both orthologues and paralogues of PERs are highly conserved, but discovering correlations between sequence differences and their functional diversity has proven difficult. By combining meta-analysis of transcriptomic data and sequence alignments, we discovered a correlation between three key amino acid positions and gene expression in response to biotic and abiotic stresses. Phylogenetic analysis revealed evolutionary pressure on these amino acids toward stress responsiveness. Using AlphaFold modeling, we found unique interdomain and protein-heme interactions involving those key amino acids in stress-induced PERs. Plausibly, these structural interactions may act as “gate keepers” by preventing larger substrates from accessing the heme and thereby shifting PER function from consumption to the production of ROS.

Keywords: AlphaFold; Arabidopsis; class III peroxidases; plant stress response; phosphate deficiency; ROS signaling



Citation: New, J.; Barsky, D.; Uhde-Stone, C. ROS Consumers or Producers? Interpreting Transcriptomic Data by AlphaFold Modeling Provides Insights into Class III Peroxidase Functions in Response to Biotic and Abiotic Stresses. *Int. J. Mol. Sci.* **2023**, *24*, 8297. <https://doi.org/10.3390/ijms24098297>

Academic Editor: Irmgard Tegeder

Received: 15 March 2023

Revised: 2 May 2023

Accepted: 4 May 2023

Published: 5 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Plants are increasingly exposed to both biotic and abiotic stresses [1], a problem exacerbated by climate change. To prevent global yield losses, a better understanding of the mechanisms by which plants cope with multiple stresses is needed. Class III peroxidases (EC 1.11.1.7, PERs) are plant-specific heme oxidoreductases at the crossroads of biotic and abiotic stress responses and have been put forward as promising candidates for crop improvement [2]. In plants, PERs have evolved into large gene families, with over 150 members in the bryophyte *Marchantia*, 138 in rice, and 73 in *Arabidopsis* [3].

Such a large number of genes, combined with wide variations in their spatiotemporal expression, suggests functional specializations [4–6]. Indeed, many substrates of PERs have been identified, including lignin precursors, auxin, and secondary metabolites [4–6]. Prior research has shown that PERs function in cell wall metabolism, cell elongation, wound healing, auxin metabolism, seed germination, defense against pathogens, and signaling in response to biotic and abiotic stresses [5,6]. However, linking these functions to specific PERs has proven difficult, in part because, in vitro, these peroxidases show low substrate specificity.

Most PERs are secreted into the apoplast, where they are thought to modulate levels of apoplastic ROS (reactive oxygen species) both positively and negatively [4,6]. The traditional view of peroxidases holds that they oxidize a substrate, such as a lignin precursor

while reducing H_2O_2 to H_2O . However, PERs can also produce ROS: hydroxyl radicals (OH^\bullet) in the hydroxylic cycle and superoxide anions ($\text{O}_2^{\bullet-}$) in the oxidative cycle. The superoxide anion is immediately converted to H_2O_2 and O_2 either spontaneously or through superoxide dismutase [7]. This generation of free radicals can have various and sometimes contrasting functions, such as cell wall stiffening by promoting covalent bonds or cell wall loosening due to the breakdown of polysaccharides [6,8,9].

Interestingly, PERs function in both stress reception and responses, and specific PER expressions are induced by various biotic and abiotic stresses, such as pathogens, metals, ozone, temperature, anoxia, and nutrient deficiencies [5,10]. PERs are also known to play a central role in plant defense both by cell wall cross-linking and by producing large amounts of toxic ROS in an oxidative burst—part of the hypersensitive response in plants [5,11,12]. In addition, ROS are known to act in signal transduction pathways in defense against pathogens [2,5,12]. Evidence is accumulating that ROS-producing PERs are also involved in the signaling of abiotic stress, such as metal toxicity and phosphate (P_i) deficiency [2,4,5,13]. We hypothesize that PERs that are upregulated in response to biotic and short-term abiotic stresses are likely producers of ROS, either in defense against pathogens or as part of a signal transduction pathway.

Even in evolutionarily distant plant families, PER protein sequences of both orthologues and paralogues are highly conserved, yet the discovery of correlations between sequence differences and their functional diversity has proved difficult [3,14]. Models of enzymatic mechanisms developed over the last century were greatly advanced with the advent of X-ray crystallographic structures in the 1990s. The first PER crystal structure solved was in peanut [15], quickly followed by a horseradish PER [16]. Of the 73 PER paralogues in Arabidopsis, only the structures of PER53 (alternative name ATPA2) [17] and PER59 (alternative name peroxidase N) [18] have been crystallographically determined. As expected, the crystal structures are all very similar and confirmed earlier models [19] that show that PERs are composed of two largely α -helical domains, one “distal” and one “proximal”, each of which binds a calcium ion, as well as a third, smaller “ β -domain”. The structure is stabilized by a network of polar and non-polar interactions orienting the heme, four to five conserved intra-domain disulfide bridges, and the two Ca^{2+} binding sites [20–24].

Despite the available structures, previous modeling approaches did not reveal clearly distinguishable structural features among PERs that linked them causally to the production or consumption of ROS [5,25]. With the emergence of AlphaFold, a recent breakthrough in protein structure modeling [26–28], we saw a much-improved opportunity to analyze structural differences that might be responsible for some of the functional diversity among PERs.

In this paper, we combine transcriptomic meta-analysis, bioinformatics—including multiple sequence alignment, motif search, and phylogenetic analysis—and AlphaFold modeling to identify key differences within the Arabidopsis PER family that potentially have structural/functional consequences related to ROS production or consumption.

2. Results

2.1. Transcriptomic Meta-Analysis Reveals Similarity of Gene Expression Patterns between Biotic and Short-Term Abiotic Stress Responses

Because of the many experiments performed on Arabidopsis and because all 73 class III peroxidase (PER) paralogues in this model plant have been identified [21], we confined our analyses, for the most part, to an examination of the gene expression patterns of PERs in Arabidopsis.

To compare expression patterns of the 73 PER paralogues in Arabidopsis, we used Genevestigator, a curated collection of normalized and systematically annotated transcriptomics data that enables a comparison of the expressions from thousands of experiments, including a vast selection of studies involving biotic and abiotic stresses [29]. While Genevestigator contains both Affymetrix and RNA-seq data, we found Affymetrix data better suited for cross-experimental comparisons, likely because of standardized protocols and

the text. (a) Root response to P_i deficiency, showing an almost opposite pattern of differential gene expression in short-term (1 h, 6 h, and 24 h; top 3 rows), compared to long-term P_i deficiency (bottom row). (b) Root responses to various pathogens (biotic stresses). Biotic stresses include infection with *Sclerotinia sclerotiorum*, *Alternaria brassicicola*, *Plectosphaerella cucumerina*, *Phytophthora infestans*, *Pseudomonas syringae*, *Liriomyza huidobrensis*, and *Hyaloperonospora arabidopsidis*. (c) Root responses to salt stress. (d) Venn diagrams visualizing shared upregulated and downregulated genes in response to biotic stresses, short-term P_i deficiency, and salt stress.

Our Genevestigator-powered meta-analysis, summarized in Venn diagrams (Figure 1d), revealed 13 *PERs* upregulated in response to all three stresses (short-term P_i deficiency, salt, and biotic stress). In contrast, 16 *PERs* were downregulated both in short-term P_i deficiency and salt stress. No *PERs* were downregulated under all three stresses.

2.2. Bioinformatic Analysis Reveals Three Key Amino Acid Positions That Determine Membership in Either of Two Groups of *PERs*: Those That Consume and Those That Produce H_2O_2/ROS

To look for key differences between the peroxidases that are upregulated versus those downregulated in response to biotic and short-term abiotic stress, we created a multiple sequence alignment (MSA) with amino acid sequences of the 13 stress-upregulated and the 16 stress-downregulated *PERs* (Figure 2a,b).

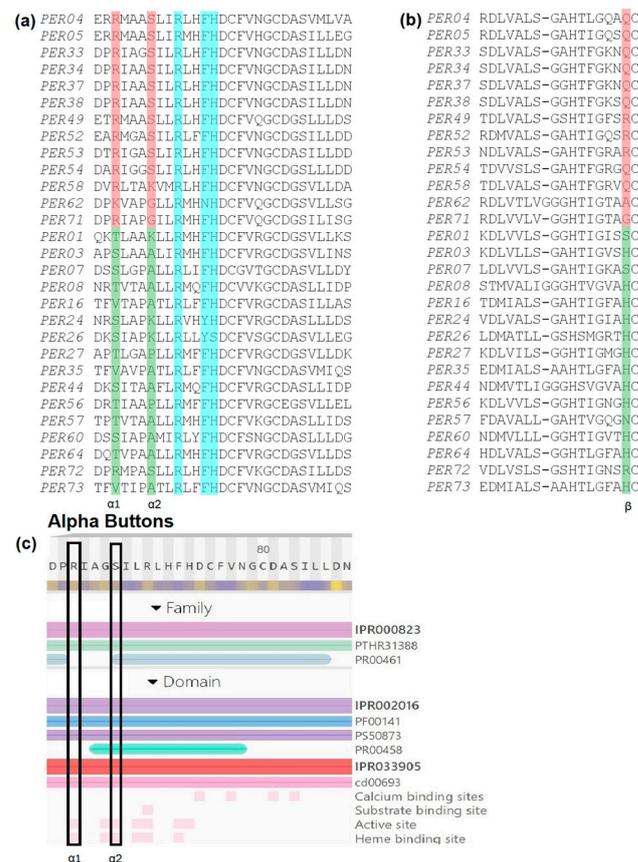


Figure 2. (a) MSA of common stress-upregulated (top 13 sequences) and downregulated (bottom 16 sequences) *PER* amino acid sequences reveals a difference between both groups at the two *Alpha Buttons*. The amino acid position is shaded in red for upregulated and green for downregulated *PER* genes. In cyan, three highly conserved active site residues are also indicated. (b) The same MSA revealed another conserved difference at the end of the first of two beta strands, at a position we termed the *Beta Button*, shaded as for the *Alpha Buttons*. (c) Support for the significance of the two *Alpha Buttons*: a position-specific scoring matrix, part of an InterProScan search, revealed both *Alpha Button* positions (marked by a black rectangle) as likely having a role in the active site and in heme binding.

Examining the MSA, we found several amino acid sites that were conserved within each group (up- or downregulated) but differed between both groups. Specifically, and significantly for our analysis, we identified three key positions that we dubbed *Alpha1*, *Alpha2*, and *Beta Buttons*. At the *Alpha1 Button* (Figure 2a), most stress-induced peroxidases contain an arginine (R), while most downregulated peroxidases contain either a threonine (T), serine (S), or valine (V). At the *Alpha2 Button* (Figure 2a), most stress-induced PERs have an S, while most downregulated PERs have an alanine (A). Finally, in the *Beta Button* position, the upregulated PERs usually have an R or glutamine (Q), while the downregulated PERs typically have a histidine (H) in that position (Figure 2b). Expanding our analysis to all 73 PERs, we found that 28 peroxidases have either R, Q, or lysine (K) in the *Beta Button* position, while 33 peroxidases contain a H in this position.

In addition to the MSA, we used InterProScan to identify functionally significant sites in these PERs. InterPro combines protein signatures from 13 databases, including the Conserved Domain Database (CDD) at NCBI, which uses a position-specific scoring matrix (PSSM). In this case, the CDD results provided additional evidence that both the *Alpha Buttons* are significant and may play a role in both the active site and heme binding (Figure 2c).

In summary, we found that stress-induced PERs tend to contain R or K at the *Alpha1 Button*, S at the *Alpha2 Button*, and R or Q at the *Beta Button*. Conversely, downregulated PERs tend to contain T, S, or V at the *Alpha1 Button*, A or K at the *Alpha2 Button*, and H at the *Beta Button*. The *Alpha1* and the *Beta Buttons* for each PER are listed in Figure 1 across the top of each profile (for simplicity, only these two out of the three “buttons” are shown), displaying, along with any exceptions, the strong correlation of these buttons with gene expression patterns. In the next section (Section 2.3), we will see that there also appears to be a strong phylogenetic correlation with PER stress-response regulation and, moreover, with the three buttons identified above. Lastly, in Section 2.4, we will see that analysis of the 3D structure allowed us finally to infer the functional significance of these buttons.

2.3. Phylogeny Reveals Correlation between Gene Expression and Evolutionary Relationship

As mentioned, the level of conservation among the multigene family of Arabidopsis class III peroxidases (PER) is quite high; by pairwise alignment, we found roughly 40% to 70% sequence identity in PER sequences—even higher among recently duplicated genes. (For example, the amino acid sequences of the mature proteins PER33 and PER34 are 94% identical and in fact cannot be distinguished in microarray analysis.) In order to look for possible correlations between observed expression differences and phylogeny, we generated a maximum likelihood tree of all 73 PERs. We based this tree on an MSA of RNA-coding sequences rather than amino acid sequences because the high conservation of PER proteins made an amino acid-based MSA less robust. To assess the reliability of nodes, we used the aLRT (approximate likelihood-ratio test) [31] (Figure 3; rectangular version available in the Supplementary Material as Figure S2), which provided higher confidence scores for the larger clades than traditional bootstrapping.

The resulting tree (Figure 3) contains four large clades (clades 1, 2, 3, and 4). Interestingly, we found a strong phylogenetic correlation with gene expression in response to short-term P_i deficiency in clades 3 and 4: most members of clade 3 (Figure 3, magenta) are upregulated in short-term P_i deficiency, while most members of clade 4 are downregulated (Figure 3, blue). Clades 1 and 2 include both upregulated and downregulated genes.

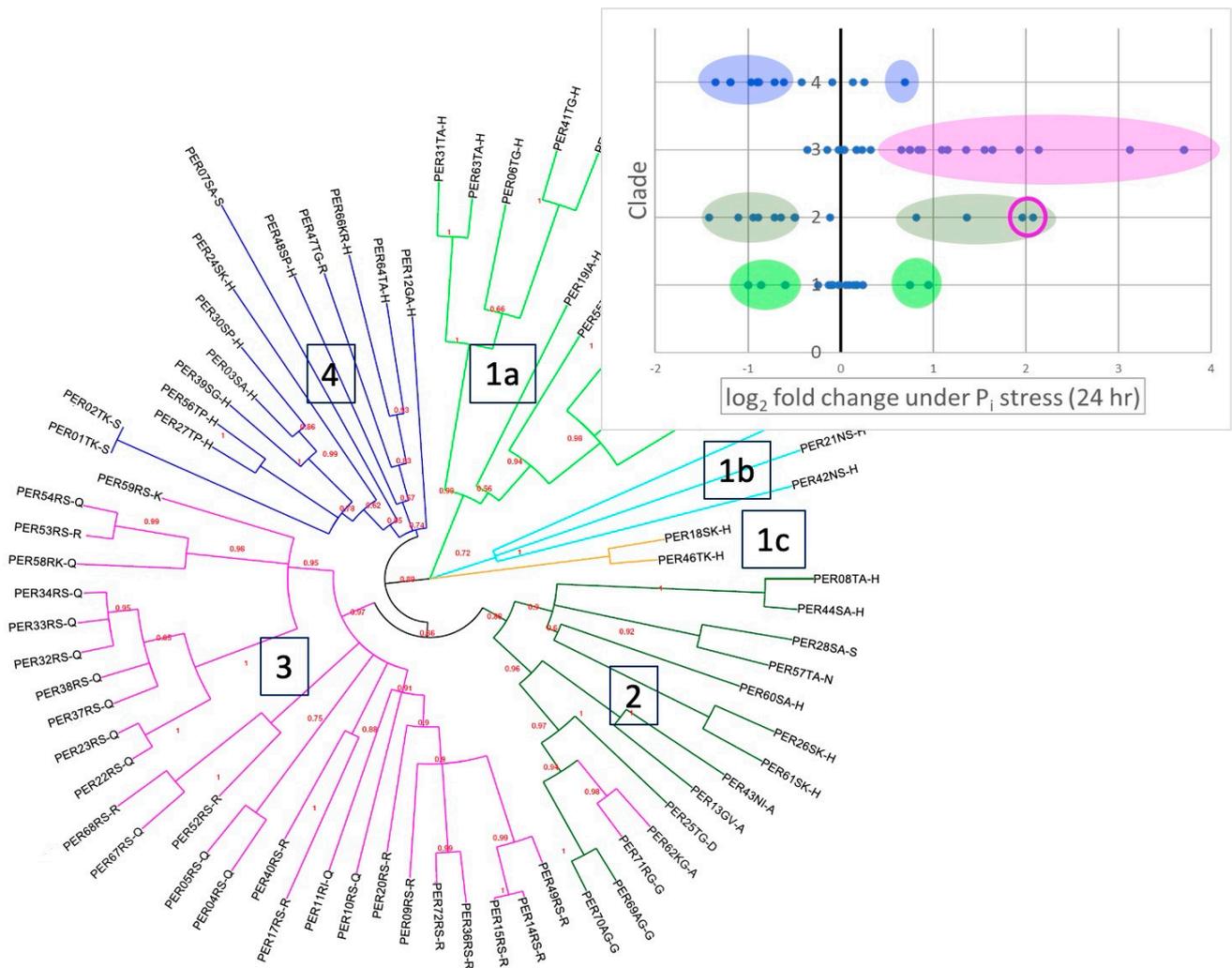


Figure 3. Maximum likelihood tree based on the 73 PER coding RNA sequences. The four main clades are distinguished by color and labeled 1–4. Approximate likelihood-ratio values are shown in red near the corresponding nodes. The particular amino acid at the *Alpha1* and *Alpha2 Buttons* and the *Beta Button* are noted at the end of each protein name, using the single letter code. Note that Supplementary Figure S2 shows all three buttons in a rectangular phylogenetic tree for better legibility. The insert shows gene expression as \log_2 fold change under short-term (24 h) P_i deficiency, with $\pm 0.5 \log_2$ fold differentially expressed genes, highlighted by color shading. Circled in magenta are PER62 and PER71, which we postulate have convergently evolved to function like members of clade 3.

While the pattern of *Alpha* and *Beta Buttons* described in Section 2.2 generally also correlate with the phylogenetic clades, there are two interesting exceptions: PER71 and PER62 belong to clade 2 yet have residues in the *Alpha1 Button* that are typical for clade 3. Moreover, in contrast to other members of clade 2, the expression pattern of PER71 and PER62 resembles that of clade 3, with both members being highly upregulated in response to short-term P_i deficiency and biotic stresses (circled in the inset graph of Figure 3).

2.4. Crystal Structures and AlphaFold Modeling Reveals the Alpha and Beta Buttons as Possible Gate Keepers That Modulate Access to the Heme

To identify possible structural implications of the *Alpha1*, *Alpha2*, and *Beta Buttons*, we sought to put the newly available 3D structural models of PERs created through AlphaFold together with the crystal structures of PER53 [17] and PER59 [18]. According to AlphaFold, the confidence levels of the models are overall very high (shown in

Supplementary Figure S3), except for the N-terminal signal peptides, which are not part of the mature protein.

To make detailed comparisons, we used the PER53 crystal structure as a reference and performed structural pairwise alignments of PER models to PER53 (PDB code: 1pa2) using the LGA program [32,33]. To put these structural alignments in perspective, we also aligned the AlphaFold model of PER53 as well as the crystal structure of PER59. All the resulting alignments—each pairwise to PER53 (1pa2)—were achieved with nearly all residues included and at low overall RMSD (root-mean-square deviation, listed in Table 1). A true prediction (not simply a copy of experimental data), the AlphaFold model of PER53 nevertheless fit the crystal structure of PER53 very tightly with an RMSD of 0.25 Å, meaning that the average backbone deviation was only a quarter the diameter of a hydrogen atom. More astonishing, however, is that the PER33 model fits the crystal structure of PER53 also very tightly, slightly more tightly even than the crystal structure of PER59!

Table 1. The pairwise alignment of the PER models created by AlphaFold, as well as the PER59 crystal structure, to the PER53 crystal structure. In each case, the reported RMSD is a measure of the overall deviation of the two backbones based on the corresponding C α atoms of aligned residues (see Section 4).

| Model | RMSD (Å) | Residues Aligned |
|------------------------|----------|------------------|
| PER01 | 1.42 | 295 |
| PER33 | 0.85 | 304 |
| PER44 | 1.25 | 289 |
| PER53 (AlphaFold) | 0.25 | 303 |
| PER59 (PDB code: 1qgj) | 1.05 | 297 |

The crystal structures also contain a bound heme group (“heme b” or protoporphyrin IX), which is required for the functioning of all peroxidases (Figure 4a). Given the very high structural similarity of the PER models to the PER53 crystal structure, we included the heme in our PER models by simply copying the atom coordinates from the PER53 structure after the alignment. This way of modeling the bound heme worked exceedingly well as judged by only a single, near clash in PER44. (The gamma carbon of an isoleucine side chain (I183) was positioned within 1.4 Angstroms of a terminal carbon atom on the α edge of the heme. This near clash is of no consequence as it could have been relieved by a tiny adjustment of the isoleucine side chain or by a fleeting application of molecular dynamics. No other clashes were reported in any of the models examined here.) Protruding from the fourth (γ) edge of the heme are two negatively charged propionate groups. Significantly, studies have suggested that compounds oxidized by PER (such as lignin precursors) require access to a heme edge or to the propionate groups [34].

As mentioned, the PER structure entails a “distal” and a “proximal” domain, as well as a third, smaller “ β -domain” (Figure 4a). A useful analogy is that the overall protein structure resembles a mouth, where the heme plays the role of the tongue and where the β -domain functions as a lower lip that can meet the propionate tips of the tongue.

Using the crystal structures and the AlphaFold models, we found that, although the *Alpha Buttons* are both separated by about 150 amino acids from the *Beta Button* in their primary structures, in their tertiary structures, all three buttons are in close proximity and at the opening of the “mouth” of the active site. Structural analysis allows us to propose a similar role for all of the three buttons, resulting in either a “closed mouth” or an “open mouth” conformation. In particular, by analyzing members of both groups, we found that peroxidases with either an R or Q in the *Beta Button* tend to form a hydrogen bond with the near heme propionate group (on the left in Figure 4a, front view, shown in the Supplementary Material), leading to a “closed mouth” conformation (Figure 4b, left panels), while the peroxidases with an H in the *Beta Button* do not form such bonds, because the amino acid is too short to reach the heme propionate tips, leading to an “open mouth” conformation (Figure 4b, right panels). A similar analysis of the *Alpha Buttons* leads to a

similar conclusion. In summary, the bigger polar side chains R, Q, and K of the *Alpha1* and *Beta Buttons* and the smaller polar S at the *Alpha 2 Button* are able to form hydrogen bonds that “button up” the entrance to the substrate binding pocket, and thus likely restrict access of larger substrates to the heme and its propionate groups. Stereo views of the hydrogen bonding details described here are presented in the Supplementary Materials (Figure S4). These results are summarized in Table 2.

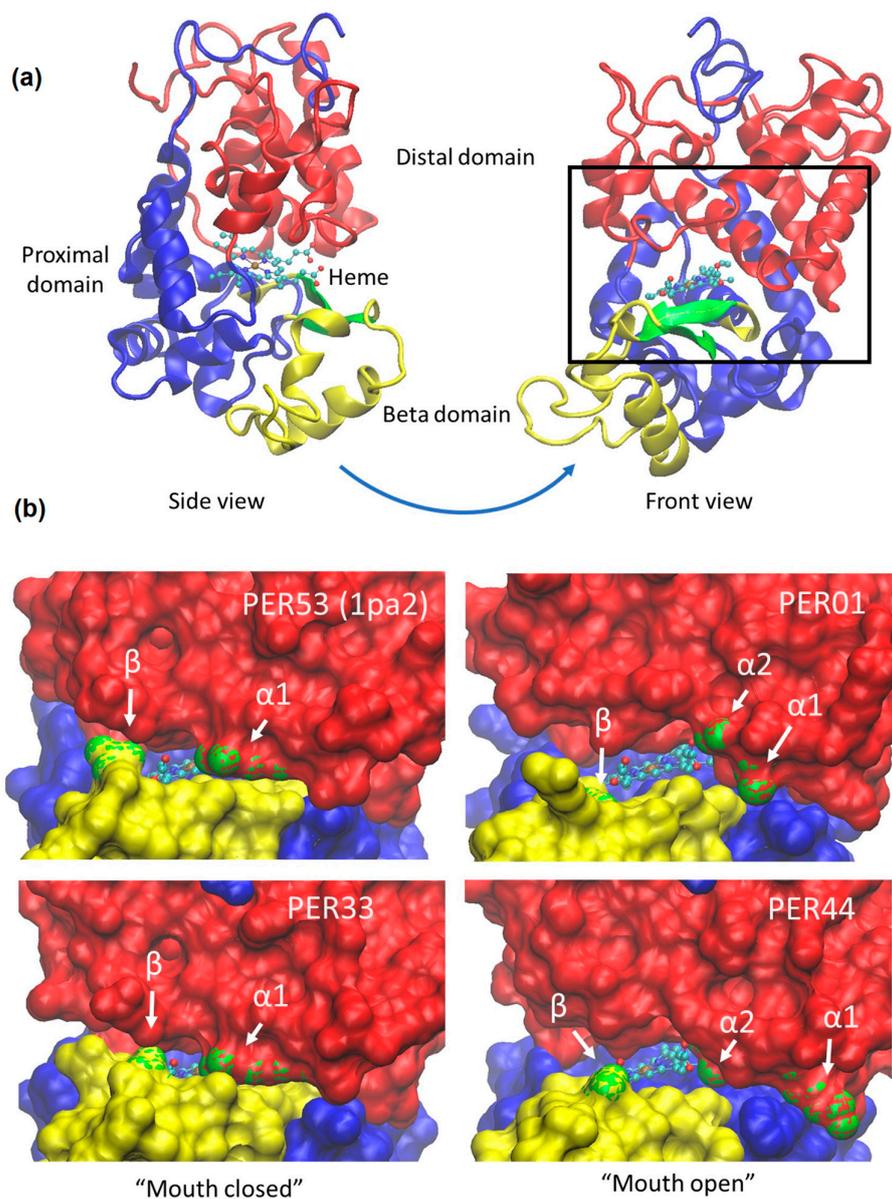


Figure 4. (a) Cartoon view of PER53 based on the crystal structure [17], with bound heme (ball and stick model) located between the distal domain (red), the proximal domain (blue), and the beta domain (yellow and green). The heme propionate groups, each ending with a pair of oxygen atoms (red balls), emerge from the γ -edge of the heme (to the right in the side view, toward the viewer in the front view). (b) “Front” views into the “mouth” of PER AlphaFold models and PER53 crystal structure with coloring and orientation as in Figure 4a (black rectangle), except here the green highlights are *Alpha1* ($\alpha 1$), *Alpha2* ($\alpha 2$) and *Beta* (β) Button residues (indicated by arrows). In the left column (PER53 and PER33,) *Alpha2* is occluded from view, and the mouth is less open. In the right column (PER01 and PER44), the mouth is more open, as evidenced by more of the bound heme being visible. We hypothesize that the less open mouth restricts access to the heme for larger substrates while allowing access for small molecules such as H_2O_2 , H_2O , $\bullet\text{OH}$, $\text{O}_2\bullet^-$, etc.

Table 2. A summary of the frequency of amino acids at each of the three buttons, along with the observed correlation with regulation and mouth posture/heme access. Next to the single-letter amino acid codes are the number of PERs that contain that amino acid in that position, culled from our MSA (see Methods and Supplementary Figure S5). Amino acids that appear most often are in bold; those in only one or two PERs are in lowercase.

| Stress-Regulation | Mouth Prediction | Alpha1 Button | Alpha2 Button | Beta Button |
|-------------------|------------------|---------------------------------------------|-------------------------------------|----------------------------------|
| Upregulated | Closed | R (12), k | S (10), g, k | Q (8), R (3), a, g |
| Downregulated | Open | S (6), T (6), V (3), r | A (10), K (3) *, p, s | H (12), s, n, r |

* Although lysine (K) is a polar residue, it apparently cannot make a hydrogen bond with the heme because (in contrast to serine) the lysine side chain is too long, given the position of the protein backbone relative to the heme. We, in fact, observe this in our model of PER01 (see Supplementary Figure S4).

3. Discussion

What distinguishes the many PER paralogues from each other, and what causes some PERs to consume and others to produce ROS? To address these questions, we started here with an exploration of the expression patterns of Arabidopsis class III peroxidase genes (*PERs*) in roots during cellular responses to biotic and short-term abiotic stress, specifically P_i deficiency and salt stress. Several of the genes that we have identified here as upregulated have been shown in prior research to be induced in response to pathogen attack and P_i deficiency, namely *PER33/34* (too similar to be distinguished in the microarray analysis), *PER62*, and *PER71* (reviewed in [5]).

We have been particularly interested in the contrasting functions of H_2O_2 -consuming and ROS-producing peroxidases. These differences may be caused by variations in reaction conditions, particularly the availability of suitable substrates, or by structural differences in the enzymes. It seems likely that *PER* genes that are induced by stress, especially by biotic stresses, tend to be producers of ROS, as ROS production is a well-documented response to pathogen and insect attacks. Such correlation is further supported by reverse genetics revealing *PER33/34* (induced by biotic and abiotic stresses) as ROS producers [35]. Comparing up- versus downregulated *PERs* by MSA, we found a strong correlation between gene expression and the three key amino acid differences, which we have termed the *Alpha* and *Beta Buttons*.

Using AlphaFold modeling, we discovered evidence of structural variations that correlate with *PER* expression patterns in response to biotic and abiotic stresses and can be understood through the structural actions of the buttons. The AlphaFold models specifically revealed that the group of stress-induced *PERs* (likely ROS producers) tend to have hydrogen bonds involving the *Alpha* and *Beta Buttons* that “button up” the entry to the substrate access channel, which could prevent access by larger substrates to the heme, preventing their oxidation. This conclusion is supported by the finding that *PER33*, *PER34*, and separately *PER53*—all “closed mouth” *PERs*—have been shown to produce H_2O_2 bursts in response to biotic stress [36,37]. On the other hand, *PERs* that are not stress-induced (likely H_2O_2 consumers) tend not to form these hydrogen bonds, thereby allowing larger substrates access to the heme. These findings are in agreement with early thoughts on the subject by Schuller et al. (1996), who suggested that the differential functioning of class III peroxidases may be caused by changes in the access to the active binding site through the substrate access channel [15].

The crystal structure of the putative *PER53* orthologue in switchgrass revealed several protein–heme hydrogen bonds involving six amino acids [24]; three of those correspond to the *Alpha* and *Beta Buttons* identified here. Although the other three may turn out to also play a role in *PER* function in terms of ROS production or consumption, they did not stand out in our analysis as being strongly correlated with expression or clade. Ironically, in addition to *PER53*, the only other crystal structure of Arabidopsis *PERs* is *PER59* (N) which is also a “closed mouth” *PER*. It is worth emphasizing that this has meant that, without the AlphaFold models, we would not have been able to observe any “open mouth” isoforms.

Furthermore, the apparent accuracy of the AlphaFold models combined with the high structural conservation of the PER family has allowed detailed analysis of the role of the key identified residues. In addition to the importance of the three buttons identified and the modeling detail at the level of hydrogen bonds, we want to also point out that the open vs. closed mouth conformation that is shown in Figure 4b appears to extend somewhat beyond the side chains. In Table 1, we tabulated the tightness of the structural alignments, and while all the RMSD values are reassuringly low, we also notice that on the one hand of the three AlphaFold models, PER33 is the closest to PER53, which makes sense given that PER33 and PER53 are, by our analysis, both “closed mouth” isoforms. On the other hand, also as one would expect, the two least tightly fitting models are both “open mouth” isoforms: PER01 and PER44.

Unfortunately, there is a dearth of data about which PERs produce or consume ROS. As mentioned above, PER33, PER34, and PER52 have been shown to produce ROS [5,36,37] and are in support of our model. PER57, however, which showed downregulation in both early P_i and salt stress in our meta-analysis, has been reported to produce ROS [38], contradicting the strong correlation with regulation. PER71, an interesting case of possible convergent evolution, has been described as a likely producer of H_2O_2 /ROS, as revealed by reverse genetics [39], but it has also been described as an enzyme involved in lignin production, a reaction that consumes H_2O_2 [6]. This possibly dual nature may indicate that some PER may at different times act in more than one catalytic cycle, depending possibly on environmental factors, particularly the availability of a suitable substrate.

Despite these possible exceptions, the key residues that we have identified in the Arabidopsis PER family on the basis of expression profiles were shown here to have plausible roles in buttoning up or leaving open the enzyme, i.e., allowing or restricting access of large molecules to the heme. Broadly, it appears that the tendency of a PER to be closed mouth versus open mouth is coded into the sequence of the individual PERs and that this, in turn, controls H_2O_2 consumption versus ROS production. Forthcoming biochemical experimentation, as well as crystal structures of consumer PERs, should further illuminate this.

4. Materials and Methods

4.1. Meta-Analysis of Gene Expression

Beginning on 11 August 2021, we employed Genevestigator (<https://genevestigator.com>) (v9.6.1), a curated database of transcriptomic data for meta-analysis [29,30] to examine the expression patterns of Arabidopsis class III peroxidases (PERs) in response to stress. To this end, all 73 Arabidopsis PER locus numbers were obtained from TAIR (<https://www.arabidopsis.org>) and used to produce a complete gene data set within Genevestigator. We then selected the Affymetrix platform (ATH1) data sets and focused on experiments of plant roots under various biotic and abiotic stresses.

Based on our previous research and literature searches, we were especially interested in experiments that showed differential gene expression in response to abiotic stresses in roots. Comparing the many available experiments in response to abiotic stress, we noticed that short-term P_i deficiency and salt stress studies showed especially high differential gene expression among PERs. Based on this observation, we selected the P_i deficiency experiments At-00524 and At-00519, which were specific to roots and included early (1 h, 6 h, and 24 h) and late (10 days) time periods of P_i deficiency. For salt stress, we selected At-00656, which was also specific to roots and had similar time points (1 h to 48 h) compared to the early P_i deficiency experiments. For biotic stress, we selected a sample of eight experiments that had similar time points (6 h to 96 h) to the early P_i deficiency and salt stress studies (At-00106, At-00108, At-00393, At-00553, At-00638, At-00661, At-00672, At-00681).

4.2. Multiple Sequence Alignment (MSA) and Site-Specific Analysis to Identify Key Amino Acids

We used the MSA editor Jalview (v2.11.2.5) [40] to edit the amino acid sequences of all 73 PERs. We trimmed the signal peptides and the variable tails from all 73 class III AtPERs and removed one large insert in a single sequence (PER48). The edited FASTA sequences are provided as supplementary data in the Supplementary File.

To identify key amino acids correlated to gene expression, we used MUSCLE [41] with default settings for protein alignment, employing a BLOSUM62 matrix to perform an MSA of the amino acid sequences of the 13 PERs upregulated under all three stress conditions (P_i deficiency, salt, and biotic stress), and the 16 PERs downregulated under both P_i deficiency and salt stress (Figure 1d) [41].

Finally, we sorted all 73 AtPER amino acid sequences so that sequences with an R in the *Alpha1 Button* were at the top of the list, then used MUSCLE (v3.8.31) to produce an MSA of all 73 PERs (Figure S5).

To further identify positions of functional importance, we submitted a representative sample of those 29 PER sequences to InterPro (v87.0) (<https://www.ebi.ac.uk/interpro>), scanning for protein signatures from 13 databases, including the Conserved Domain Database (CDD) at NCBI, Panther, Pfam, and Prosite [42].

4.3. Phylogeny

We used Phylogeny.fr (<http://www.phylogeny.fr/index.cgi>) to run alignments in the system's "A la carte" mode [31]. We employed MUSCLE again to generate an MSA and to produce a maximum likelihood phylogenetic tree [43]. In this case, because of the high conservation of PER proteins, we used RNA-coding sequences from the 73 PERs rather than amino acid sequences. We constructed a phylogenetic tree using the PhyML program (v3.1/3.0 aLRT). Reliability for internal branches was assessed using the aLRT test (SH-Like) [44]. We collapsed branches with support values smaller than 50%. The visualization of the phylogenetic tree was produced using TreeDyn (v198.3) [45].

4.4. Structural Analysis of Crystal Structures and AlphaFold 3D Protein Models

To identify the structural roles of the *Alpha* and *Beta Button* residues, we downloaded models from the AlphaFold Protein Structure Database (v2.1.1) (<https://alphafold.ebi.ac.uk>) [28], which has produced models of a vast number of proteins using the AlphaFold2 program [27]. Choosing several representatives from the upregulated and downregulated groups, we examined the structural models near the "mouth", particularly access to the tips of the "tongue", which are formed by the propionate groups of the bound heme.

We performed structural alignments and analysis of the models as follows:

- (1) We removed the N-terminal signal peptide regions from the downloaded models (typically 20 to 40 residues). N-terminal signal peptides are removed in the formation of the mature (secreted) protein but were nevertheless present in all the models. In all the models inspected, these peptides protruded unnaturally and were visually removed, and these identifications were additionally supported by the signal peptide program SignalP (v6.0) [46].
- (2) The models were pairwise structurally aligned to the crystal structure of PER53 (PDB code: 1pa2) [17], using LGA [32,33]. The alignments were obtained through "local-global" structural comparisons without a preassigned residue correspondence (option 4) and with the default distance cutoff of 5 Å. The coordinates of the heme molecule were simply copied over to the model from the 1pa2 coordinate file after the structural alignment of the model.
- (3) The aligned models were analyzed using the VMD program [47], which enabled us to check for clashes, identify hydrogen bonds (protein-protein and between protein and heme), as well as to visualize the mouth and heme accessibility (see Figure 4 and Supplementary Figure S4).

- (4) In order to make the correct correspondence of the buttons in each PER, we used the following simple motifs based on our MSA (see Supplementary Materials). These motifs are valid for the entire Arabidopsis PER family:
- Alpha1* and *Alpha2*: 1xxx2xxRxxfhDC, where 1 and 2 correspond to *Alpha1* and *Alpha2*, respectively, and RxxfhDC is the highly conserved active site. Uppercase letters are 100% conserved, lowercase letters are mostly conserved, and “x” is variable.
 - Beta: HtxGxxBCxxxxxR, where B corresponds to the *Beta Button*.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms24098297/s1>.

Author Contributions: Conceptualization, J.N., D.B. and C.U.-S.; methodology, J.N., D.B. and C.U.-S.; formal analysis, J.N., D.B. and C.U.-S.; writing J.N., D.B. and C.U.-S.; visualization, J.N., D.B. and C.U.-S.; funding acquisition, C.U.-S. All authors have read and agreed to the published version of the manuscript.

Funding: The APC and release time for CUS was funded by an internal 2022–2023 RSCA Grant, CSUEB.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Pascual, L.S.; Segarra-Medina, C.; Gomez-Cadenas, A.; Lopez-Climent, M.F.; Vives-Peris, V.; Zandalinas, S.I. Climate change-associated multifactorial stress combination: A present challenge for our ecosystems. *J. Plant Physiol.* **2022**, *276*, 153764. [[CrossRef](#)] [[PubMed](#)]
- Kidwai, M.; Ahmad, I.Z.; Chakrabarty, D. Class III peroxidase: An indispensable enzyme for biotic/abiotic stress tolerance and a potent candidate for crop improvement. *Plant Cell Rep.* **2020**, *39*, 1381–1393. [[CrossRef](#)] [[PubMed](#)]
- Mbadinga Mbadinga, D.L.; Li, Q.; Ranocha, P.; Martinez, Y.; Dunand, C. Global analysis of non-animal peroxidases provides insights into the evolution of this gene family in the green lineage. *J. Exp. Bot.* **2020**, *71*, 3350–3360. [[CrossRef](#)]
- Passardi, F.; Cosio, C.; Penel, C.; Dunand, C. Peroxidases have more functions than a Swiss army knife. *Plant Cell Rep.* **2005**, *24*, 255–265. [[CrossRef](#)] [[PubMed](#)]
- Cosio, C.; Dunand, C. Specific functions of individual class III peroxidase genes. *J. Exp. Bot.* **2009**, *60*, 391–408. [[CrossRef](#)]
- Shigeto, J.; Tsutsumi, Y. Diverse functions and reactions of class III peroxidases. *New Phytol.* **2016**, *209*, 1395–1402. [[CrossRef](#)]
- Noctor, G.; Reichheld, J.-P.; Foyer, C.H. ROS-related redox regulation and signaling in plants. *Semin. Cell Dev. Biol.* **2018**, *80*, 3–12. [[CrossRef](#)]
- Meng, G.; Fan, W.Y.; Rasmussen, S.K. Characterisation of the class III peroxidase gene family in carrot taproots and its role in anthocyanin and lignin accumulation. *Plant Physiol. Biochem.* **2021**, *167*, 245–256. [[CrossRef](#)]
- Passardi, F.; Penel, C.; Dunand, C. Performing the paradoxical: How plant peroxidases modify the cell wall. *Trends Plant Sci.* **2004**, *9*, 534–540. [[CrossRef](#)]
- Eljebbawi, A.; Savelli, B.; Libourel, C.; Estevez, J.M.; Dunand, C. Class III Peroxidases in Response to Multiple Abiotic Stresses in Arabidopsis thaliana Pyrenean Populations. *Int. J. Mol. Sci.* **2022**, *23*, 3960. [[CrossRef](#)]
- Almagro, L.; Gomez Ros, L.V.; Belchi-Navarro, S.; Bru, R.; Ros Barcelo, A.; Pedreno, M.A. Class III peroxidases in plant defence reactions. *J. Exp. Bot.* **2009**, *60*, 377–390. [[CrossRef](#)] [[PubMed](#)]
- Minibayeva, F.; Beckett, R.P.; Kranner, I. Roles of apoplastic peroxidases in plant response to wounding. *Phytochemistry* **2015**, *112*, 122–129. [[CrossRef](#)]
- Shin, R.; Berg, R.; Schachtman, D. Reactive oxygen species and root hairs in Arabidopsis root response to nitrogen, phosphorus and potassium deficiency. *Plant Cell Physiol.* **2005**, *46*, 1350–1357. [[CrossRef](#)]
- Duroux, L.; Welinder, K.G. The peroxidase gene family in plants: A phylogenetic overview. *J. Mol. Evol.* **2003**, *57*, 397–407. [[CrossRef](#)]
- Schuller, D.J.; Ban, N.; Huystee, R.B.; McPherson, A.; Poulos, T.L. The crystal structure of peanut peroxidase. *Structure* **1996**, *4*, 311–321. [[CrossRef](#)]
- Gajhede, M.; Schuller, D.J.; Henriksen, A.; Smith, A.T.; Poulos, T.L. Crystal structure of horseradish peroxidase C at 2.15 Å resolution. *Nat. Struct. Biol.* **1997**, *4*, 1032–1038. [[CrossRef](#)]
- Ostergaard, L.; Teilum, K.; Mirza, O.; Mattsson, O.; Petersen, M.; Welinder, K.G.; Mundy, J.; Gajhede, M.; Henriksen, A. Arabidopsis ATP A2 peroxidase. Expression and high-resolution structure of a plant peroxidase with implications for lignification. *Plant Mol. Biol.* **2000**, *44*, 231–243. [[CrossRef](#)]

18. Mirza, O.; Henriksen, A.; Ostergaard, L.; Welinder, K.G.; Gajhede, M. Arabidopsis thaliana peroxidase N: Structure of a novel neutral peroxidase. *Acta Cryst. D Biol. Cryst.* **2000**, *56*, 372–375. [[CrossRef](#)]
19. Welinder, K.G.; Gajhede, M. Structure & Evolution of Peroxidases. In *Plant Peroxidases: Biochemistry and Physiology*; Welinder, K.G., Rasmussen, S.K., Penel, C., Greppin, H., Eds.; University of Geneva: Geneva, Switzerland, 1993.
20. Singh, S.; Pandey, V.P.; Naaz, H.; Dwivedi, U.N. Phylogenetic analysis, molecular modeling, substrate-inhibitor specificity, and active site comparison of bacterial, fungal, and plant heme peroxidases. *Biotechnol. Appl. Biochem.* **2012**, *59*, 283–294. [[CrossRef](#)]
21. Welinder, K.G.; Justesen, A.F.; Kjaersgaard, I.V.; 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*, 6063–6081. [[CrossRef](#)]
22. Nielsen, K.L.; Indiani, C.; Henriksen, A.; Feis, A.; Becucci, M.; Gajhede, M.; Smulevich, G.; Welinder, K.G. 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*, 11013–11021. [[CrossRef](#)] [[PubMed](#)]
23. Gajhede, M. Plant peroxidases: Substrate complexes with mechanistic implications. *Biochem. Soc. Trans.* **2001**, *29*, 91–98. [[CrossRef](#)] [[PubMed](#)]
24. Moural, T.W.; Lewis, K.M.; Barnaba, C.; Zhu, F.; Palmer, N.A.; Sarath, G.; Scully, E.D.; Jones, J.P.; Sattler, S.E.; Kang, C. Characterization of Class III Peroxidases from Switchgrass. *Plant Physiol.* **2017**, *173*, 417–433. [[CrossRef](#)]
25. Douroupi, T.G.; Papassideri, I.S.; Stravopodis, D.J.; Margaritis, L.H. Molecular cloning and tissue-specific transcriptional regulation of the first peroxidase family member, Udp1, in stinging nettle (*Urtica dioica*). *Gene* **2005**, *362*, 57–69. [[CrossRef](#)]
26. Jumper, J.; Hassabis, D. Protein structure predictions to atomic accuracy with AlphaFold. *Nat. Methods* **2022**, *19*, 11–12. [[CrossRef](#)] [[PubMed](#)]
27. Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; et al. AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **2022**, *50*, D439–D444. [[CrossRef](#)]
28. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [[CrossRef](#)]
29. Hruz, T.; Laule, O.; Szabo, G.; Wessendorp, F.; Bleuler, S.; Oertle, L.; Widmayer, P.; Gruissem, W.; Zimmermann, P. Genevestigator v3: A reference expression database for the meta-analysis of transcriptomes. *Adv. Bioinform.* **2008**, *2008*, 420747. [[CrossRef](#)]
30. Zimmermann, P.; Hirsch-Hoffmann, M.; Hennig, L.; Gruissem, W. GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox. *Plant Physiol.* **2004**, *136*, 2621–2632. [[CrossRef](#)]
31. Dereeper, A.; Guignon, V.; Blanc, G.; Audic, S.; Buffet, S.; Chevenet, F.; Dufayard, J.F.; Guindon, S.; Lefort, V.; Lescot, M.; et al. Phylogeny.fr: Robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **2008**, *36*, W465–W469. [[CrossRef](#)]
32. Zemla, A. LGA: A method for finding 3D similarities in protein structures. *Nucleic Acids Res.* **2003**, *31*, 3370–3374. [[CrossRef](#)] [[PubMed](#)]
33. Zemla, A.; Zhou, C.E.; Slezak, T.; Kuczmarski, T.; Rama, D.; Torres, C.; Sawicka, D.; Barsky, D. AS2TS system for protein structure modeling and analysis. *Nucleic Acids Res.* **2005**, *33*, W111–W115. [[CrossRef](#)] [[PubMed](#)]
34. Kwon, H.; Moody, P.C.E.; Raven, E. Understanding the Reactivity and Interactions of Peroxidases with Substrates. In *Heme Peroxidases*; Raven, E., Dunford, B.H., Eds.; Royal Society of Chemistry: London, UK, 2015.
35. Passardi, F.; Tognolli, M.; De Meyer, M.; Penel, C.; Dunand, C. Two cell wall associated peroxidases from Arabidopsis influence root elongation. *Planta* **2006**, *223*, 965–974. [[CrossRef](#)] [[PubMed](#)]
36. Bindschedler, L.V.; Dewdney, J.; Blee, K.A.; Stone, J.M.; Asai, T.; Plotnikov, J.; Denoux, C.; Hayes, T.; Gerrish, C.; Davies, D.R.; et al. Peroxidase-dependent apoplastic oxidative burst in Arabidopsis required for pathogen resistance. *Plant J.* **2006**, *47*, 851–863. [[CrossRef](#)] [[PubMed](#)]
37. Bolwell, G.P.; Bindschedler, L.V.; Blee, K.A.; Butt, V.S.; Davies, D.R.; Gardner, S.L.; Gerrish, C.; Minibayeva, F. The apoplastic oxidative burst in response to biotic stress in plants: A three-component system. *J. Exp. Bot.* **2002**, *53*, 1367–1376.
38. Survila, M.; Davidsson, P.R.; Pennanen, V.; Kariola, T.; Broberg, M.; Sipari, N.; Heino, P.; Palva, E.T. Peroxidase-Generated Apoplastic ROS Impair Cuticle Integrity and Contribute to DAMP-Elicited Defenses. *Front. Plant Sci.* **2016**, *7*, 1945. [[CrossRef](#)]
39. Raggi, S.; Ferrarini, A.; Delledonne, M.; Dunand, C.; Ranocha, P.; De Lorenzo, G.; Cervone, F.; Ferrari, S. The Arabidopsis Class III Peroxidase AtPRX71 Negatively Regulates Growth under Physiological Conditions and in Response to Cell Wall Damage. *Plant Physiol.* **2015**, *169*, 2513–2525. [[CrossRef](#)]
40. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.; Clamp, M.; Barton, G.J. Jalview Version 2—A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **2009**, *25*, 1189–1191. [[CrossRef](#)]
41. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [[CrossRef](#)]
42. Zdobnov, E.M.; Apweiler, R. InterProScan—An integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **2001**, *17*, 847–848. [[CrossRef](#)]
43. Guindon, S.; Gascuel, O. A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Syst. Biol.* **2003**, *52*, 696–704. [[CrossRef](#)]
44. Anisimova, M.; Gascuel, O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst. Biol.* **2006**, *55*, 539–552. [[CrossRef](#)]

45. Chevenet, F.; Brun, C.; Bañuls, A.L.; Jacq, B.; Christen, R. TreeDyn: Towards dynamic graphics and annotations for analyses of trees. *BMC Bioinform.* **2006**, *7*, 439. [[CrossRef](#)]
46. Teufel, F.; Almagro Armenteros, J.J.; Johansen, A.R.; Gislason, M.H.; Pihl, S.I.; Tsirigos, K.D.; Winther, O.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat. Biotechnol.* **2022**, *40*, 1023–1025. [[CrossRef](#)]
47. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.