Previously we presented the purification, biochemical characterization, and cloning of a cationic peroxidase isoenzyme (CysPrx) from artichoke (Cynara cardunculus subsp scolymus (L.) Hegi) leaves. The protein was shown to have some interesting properties, suggesting that CysPrx could be considered as a potential candidate for industrial application. In addition, from the CysPrx sequence, two full-length cDNAs; CysPrx1 and CysPrx2, differing for three amino acids, were isolated. A three-dimensional model was predicted from CysPrx1 by homology modeling, using two different computational tools. Herein we discuss the roles of particular amino acid residues and structural motifs or regions of both deduced sequences, with the aim to find new understandings between the new plant peroxidase isoenzymes and their physiological substrates. Additionally, the obtained information may lead to new methods for improving the stability of the enzyme in several processes of biotechnological interest for peroxidase applications.

Introduction

Plant peroxidase (Prx) are coded by a large multigene family, and participate in a broad range of physiological processes including the removal of hydrogen peroxide from chloroplasts and cytosol, the oxidation of toxic compounds, the biosynthesis of cell walls (lignin and suberin), indole-3-acetic acid regulation, ethylene biosynthesis, defense responses toward wounding, and other stresses. Prxs can also catalyze the formation of diferuloyl and isodityrosine linkages in both primary and secondary cell walls. The specific functions of individual Prxs are often difficult to understand because of their low substrate specificity and the existence of many isoenzymes. The most thoroughly studied example of a higher plant Prx is horseradish (HRP), which is also the most commercially available Prx, even if one of the major challenges associated with this enzyme is its susceptibility to inactivation during thermal treatments. Soybean seed coat was also identified as a rich source of Prx. This enzyme was found to be very stable at elevated temperature and was less susceptible than HRP to permanent inactivation by hydrogen peroxide, even if catalytically slower against aqueous phenols. With the aim of finding alternative Prxs for a wide range of applications, in a previous study we examined a cationic peroxidase isoenzyme (CysPrx) from artichoke [Cynara cardunculus spp scolymus (L.) Hegi] leaves. The enzyme was purified using a combined protocol involving a two phase system, ion exchange and gel filtration chromatography. Systematic biochemical characterization of the enzyme (optimum of pH and temperature, substrate specificity) demonstrated that CysPrx possessed interesting properties, such as stability at a wide pH range, suggesting that CysPrx could be a potential candidate for application in industry requiring large pH stability. MALDI-TOF MS analyses resolved the CysPrx isoenzyme into two cationic isoforms differing by one aminoacid. The presence of these isoforms was confirmed by the isolation of full-length cDNAs encoding CysPrx that generate two slightly

Keywords: artichoke, computational studies, peroxidase, recombinant protein, cDNA sequences

Submitted: 07/06/11
Revised: 08/31/11
Accepted: 08/31/11
http://dx.doi.org/10.4161/bbug.3.1.17937
*Correspondence to: Angela Cardinali;
Email: angela.cardinali@ispa.cnr.it

Addendum to: Cardinali A, Tursi N, Ligorio A, Giuffrida MG, Napolitano L, Caliandro R, et al. Purification, biochemical characterization and cloning of a new cationic peroxidase isoenzyme from artichoke. Plant Physiol Biochem 2011; 49:395–403; PMID:21345687; http://dx.doi.org/10.1016/j.plaphy.2011.01.028
different sequences coding for two putative CysPrx: CysPrx1 and CysPrx2 sharing a very high degree of similarity with sunflower and lettuce Prx. Finally, a three-dimensional model was predicted from CysPrx1 by homology modeling, using two different computational tools. The CysPrx1 structure was predicted to follow the barley peroxidase (BP) backbone in the proximal domain and to assume the configuration of the reaction pocket typical of HRP and other active plant peroxidases, where the major role is played by the orientation of the distal His residue. In this respect, the predictions diverge from the inactive crystal structure of BP, even if BP has a higher sequence homology with CysPrx1. Peculiar properties could arise for CysPrx1 due to the presence of Met252 in its catalytic site: the presence of a sulfur atom close to the heme pocket might affect the stability of the heme group by stabilization/delocalization of the porphyrin cation of compound I and could differentiate the catalytic properties of CysPrx from those of BP and HRP.

**New Insights into the Predicted Structures**

The structure prediction was obtained from CysPrx1 by using two top performer public servers: Phyre and I-TASSER. The best models they produced, denoted respectively by modPhy and modTas, differ mainly in the terminal parts of the sequence (30 residues from the N-terminus and 20 residues from the C-terminus). The C\textsubscript{α} atoms of the two models differ by root mean square deviations (RMSD) of 0.9 Å. For modTas, an RMSD of 8.7 ± 4.5 Å from the true model may be estimated on the basis of extensive benchmark tests of the procedure.

The models may be compared with the crystal structures of the Prx having higher homology with CysPrx1, which are Barley grain (PDB code 1BGP), Arabidopsis thaliana (1PA2), horseradish (1ATJ), soybean (1FHF) and peanut (1SCH) peroxidases. As an example, RMSD of the C\textsubscript{α} atoms of modTas (squares) and modPhy (circles) from those of 1BGP are plotted in **Figure 1**, vs. the residue number. They have been calculated by the secondary-structure matching (SSM) algorithm. The shaded areas and the letters indicated the common \(\alpha\)-helix structures which have been identified by the Define Secondary Structure of Proteins (DSSP) algorithm in the modTas and modPhy models. Only the residues 30–337 have been considered for this analysis, since the remaining residues are not covered by the 1BGP crystal structure. It can be seen that the modPhy strictly follows the 1BGP backbone, apart from residues 126–130, where deviations from the 1BGP structure larger than 3 Å occur. On the other hand, modTas exhibits larger deviations from 1BGP, especially in the loop between helices B and C, comprising the residues 93–113. In this region modTas follows the common backbone features shown by all the other Prx (1PA2, 1ATJ, 1FHF and 1SCH).

The same trend is followed at the side chain level, in the region of the active site. While modPhy has a configuration of the reaction pocket very similar to that of 1BGP (Fig. 2A), where the position of

![Figure 1. RMSD between the C\textsubscript{α} atoms of the best model generated by I-TASSER (squares) and Phyre (circles) from those of the 1BGP crystal structure. The shaded areas and the letters indicate the \(\alpha\)-helices identified in the two models.](image-url)
the distal histidine (His) side chain has been addressed as the main cause of the inactive character of the enzyme, modTas follows the configuration which occurs in all the other Prx (Fig. 2B), where the position of the distal His closer to heme makes the enzymes active.

It has been argued\textsuperscript{11} that the origin of the alteration of the orientation of the distal His in the crystal structure of BP lies in a major rearrangement in the above mentioned loop region 93–113. In fact, the distal His, in its active configuration, is hydrogen bonded with a conserved distal asparagine (Asn) residue included in that loop region (residue Asn106 of CysPrx1). In the inactive configuration this characteristic hydrogen bond is disrupted and a new one is formed, which involves a threonine (Thr) in 1BGP and serine (Ser) 96 in modPhy, which also lie in the same, differently shaped loop.\textsuperscript{12}

The two predicted models have been subjected to a short (50 ps) unrestrained equilibration by using NAMD.\textsuperscript{13} The results are shown in Figure 3, where the initial and final models are superimposed for modPhy (a) and modTas (b). The most striking results regards the orientation of the distal His: it changes for modPhy, drifting toward a configuration closer to that of the active HPR, while it remains stable for modTas.

The predicted structures have now been analyzed in relation to the disulphide bridges they form. The disulphide bridges formed in modPhy and modTas are reported in Table 1. The known structure of BP has four disulphide bridges: one in the distal domain (last row of Table 1) and the others in the proximal domain. This feature is common to all the above mentioned Prx structures having higher identity with CysPrx1. However, although all the cysteine (Cys) residues forming disulphide bridge are conserved in CysPrx1, the predicted models do not reproduce this feature. Two disulphide bridges are conserved in both the predicted structures, modPhy lacks the first disulphide bridge, while modTas lacks two disulphide bridges. On the other hand, the occurrence of the expected disulphide bridges for CysPrx1 is highly probable, as has been assessed by the server disulfide bounding connectivity pattern (DBCP).\textsuperscript{14}

The calculated probabilities of occurrence are reported in the last column of Table 1. It should be noted that the nature of the discrepancies between predictions and expectations regarding the disulphide bridges is very different for the two models. For the cases where the bridges are not formed, the S-S distance has

| Table 1. Sequence numbers of cysteine residues forming disulphide bridges in 1BGP and in the models predicted by Phyre (modPhy) and I-TASSER (modTas). In the last column the probability of their occurrence calculated for the sequence CysPrx1 are reported. When the bridges are not formed, the S-S distance is reported. |
| --- |
| 1BGP | CysPrx1 | ModPhy | ModTas | Occurrence Probability (%) |
| 18–99 | 47–128 | d = 14.6 Å | d = 3.3 Å | 69 |
| 51–56 | 80–85 | yes | yes | 99 |
| 106–301 | 134–329 | yes | d = 3.2 Å | 79 |
| 186–213 | 214–241 | yes | Yes | 67 |
been measured in the models are reported in Table 1. From these values it is evident that for modTas the bridges are not formed because of wrong conformations of the Cysrotamers, while for modPhyre the cause is more fundamental, as Cys128 falls in the region of most deviation from 1BGP, consisting in the B-C loop (residues 93–113). This indicates that the hypotheses of Phyre, which predicts a model for CysPrx1 very similar to that of BP, is not consistent with the properties related to the formation of disulphide bridges, and supports the conclusion arising from the previous study of the active site equilibration.

New Structure Predictions

In order to achieve more realistic structural predictions, the possibility to include restraints in the generation of models has been exploited, by using the a priori information about the experimentally observed disulphide bridges. The Phyre server does not have this option, so it was not used for this further analysis. The I-TASSER server allows specifying two types of restraints: one regarding the distances between specific atoms, the other regarding the residues to be put in contact. They were both tried on the eight cysteines involved in the disulphide bridges (the first one was implemented by requiring a distance of 2.9 Å between the two sulfur atoms), obtaining two new models from the CysPrx1 sequence. The estimated RMSD from the true model was 8.1 ± 4.4 Å for the model built using the restraint on atomic distances, and 7.9 ± 4.4 Å for the model built by using the restraint on residues in contact (they are both lower than the RMSD estimated for the previous generated model). Since the model obtained by the second type of restraint has lower distance from the true model, it was chosen for further analysis. It was manually modified by means of the program Coot, to achieve the following:

- the rotamer of Arg74 was changed to avoid collisions with the heme group;
- two calcium ions were included by analogy with those present in 1BGP, because of their importance in maintaining the structural integrity of heme. In addition, a structure prediction was performed also for CysPrx2. The corresponding model was generated by I-TASSER with the restraint that the eight cysteines forming the disulphide bridges (which are conserved in CysPrx2) are in contact. In this case an RMSD of 8.2 ± 4.5 Å from the true model was estimated. It underwent a similar manual modification as that described above for CysPrx1.

Apart from the terminal regions, the two models (modCys1 and modCys2 hereinafter) have a very similar backbone, which follows closely that of modTas. The calculated RMSD for Cα atoms is 0.79 Å between modCys1 and modCys2, 0.89 Å between modCys1 and modTas and 0.75 Å between modCys2 and modTas. In the active site region, modCys1 and modCys2 confirms the configuration of distal His residue common to modTas and to the crystal structures of the active Prxs.

Molecular Dynamics Calculations

The CysPrx1 and CysPrx2 models were subjected to Molecular Dynamics (MD) calculations in order to investigate their stability and the effect of the mutations R121H, A243V and G349E, which distinguish the two sequences.17

The systems were prepared as follows:

- Five histidines are present in CysPrx1 and six in CysPrx2. Their protonation state was predicted by the public server H++ by considering pH 5.5, corresponding to the maximum of the activity exhibited by the enzyme.17 The results show that most of the histidines are protonated in one position, at Nε. Therefore both the simulations were performed by protonating all the histidines in the same way, at Nε.
- D, E, K and R residues were considered in their ionized form.
- The models were solvated in a box of explicit water, ensuring that the solvent shell would extend for at least 1.6 nm around them.
- A number of Cl-counter-ions (two for CysPrx1 and four for CysPrx2) were added to the systems, in order to neutralize them.
- Periodic boundary conditions were applied.

The simulations were performed as follows:

The CHARMM (PARAM22)19 force field was used for the protein and the heme group, with Beglov and Roux force fields for chlorine counter-ions20 and Marchand and Roux force fields for calcium ions.21 The TIP3P22 model was used for water molecules. Long-range electrostatic interactions were treated with the particle mesh Ewald method.23 A Fourier spacing of 0.12 nm combined with a fourth-order cubic spline interpolation was used. A 1.2 nm cutoff was used for Van der Waals interactions as well as for the real-space part of the electrostatic interactions. All bond lengths were constrained with the LINCS algorithm and the time step was set to 2 fs. NPT ensemble (t = 298 K, p = 1 bar) MD simulations were performed by coupling the systems with a Nose-Hoover thermostat24 and an Andersen-Parrinello-Rahman barostat.25,26

A common simulation protocol was followed, consisting in the following steps. The systems underwent 2000 steps of steepest-descent energy minimization. Then they were gradually heated from 0 K up to 298 K in 12 steps of 100 ps simulation. During the simulation steps, harmonic restraints were applied to the protein, to the atoms of the heme group and to the two calcium ions, by using an elastic constant of 10 kcal mol⁻¹Å⁻². After that, a 200 ps equilibration and 10 ns MD simulations were performed without restraints. The NAMD13 code was used for MD calculations. The frames collected during the final 10 ns MD simulations were analyzed using the visual molecular dynamics (VMD) package,27 with tcl scripts written for purpose.

In Figure 4 the RMSD calculated between the structural configurations obtained along the trajectory and the initial structure for CysPrx1 and CysPrx2 are reported as a function of the simulation...
They indicate that both the systems are subjected to drift during the simulation, therefore have not reached the convergence within the simulation time. This could be due to the fact that the initial configuration of side chains, being generated by homology modeling, was very far from being stable. Moreover, it results that modCys1 and modCys2 exhibit slightly different properties in their time evolution.

In Figure 5 the root mean square fluctuations (RMSF) of each residue during the simulation is plotted, with the α-helix structure identified in the modCys1 and modCys2 models shown as shaded areas. It results that, apart from the terminal regions, the fluctuations are limited (< 3 Å), and different in the two models. In particular, modCys2 has a larger flexibility than modCys1 in the loops F''–G, where the mutation A243V is located, and D'–E. In this second case the anomalous flexibility is due to the interaction with the heme group, which experiences a drift toward this loop during the simulation. This drift is not present in the simulation of modCys1. As common features, the models have the loop B–C as the region of largest fluctuations and very small fluctuations (below 1 Å) in the regions of the Ca ions (residues 83–88 and 258–262).

In view of characterizing in more details the behavior of the B–C loop during the simulations, given its proven importance in determining the active site configuration through the hydrogen bond with the distal His, the RMSD of the residues 93–113 with respect to the corresponding residues in the homologous crystal structures have been calculated throughout the simulation. This calculation has been performed in two steps: (a) each frame has been fitted with the reference crystal structure by using the residues belonging to the α-helices; (b) the RMSD of residues 93–113 (not used in the fitting) has been measured within each frame. The resulting RMSD values averaged within the trajectory are reported in Table 2, together with their standard deviations. It results that, despite the higher sequence identity, BP is not its most probable structural homolog. In fact, a model basically reproducing the 1BGP structure with the CysPrx sequence fails to reproduce a well conserved feature of the plant Prx, i.e., the formation of four disulphide-bridges. Instead the new sequence has features, like the configuration of the B–C loop, similar to those of the other active Prx.

A different stability has been predicted for the two isoforms CysPrx1 and CysPrx2, basically triggered by the mutation A243V. Definite conclusions regarding the side chain configuration of the active site

Conclusions

Computational studies performed using the sequence of a new peroxidase iso-enzyme from artichoke, which include homology modeling, equilibration and molecular dynamics, lead to interesting predictions about its structural model. It was shown that, in spite of its higher sequence identity, BP is not its most probable structural homolog. In fact, a model basically reproducing the 1BGP structure with the CysPrx sequence fails to reproduce a well conserved feature of the plant Prx, i.e., the formation of four disulphide-bridges. Instead the new sequence has features, like the configuration of the B–C loop, similar to those of the other active Prx.
cannot be drawn, due to the lack of knowledge concerning the protonation state of the histidines and the necessity to explore longer simulation times, in order to correct for wrong conformations introduced by the homology modeling.

Table 2. Sequence Identity (Seq. Id.) and RMSD values (in Å) averaged within the simulation time of the residues 93–113, belonging to the B-C loop, for modCys1 and modCys2. The crystal structures of several homologous peroxidases have been used as reference.

| Crystal Structures         | Seq.Id. (%) | modCys1    | modCys2    |
|----------------------------|-------------|------------|------------|
| Barley Grain (1BGP)        | 53          | 6.7 ± 0.3  | 6.6 ± 0.3  |
| Arabidopsis thaliana (1PA2) | 42          | 4 ± 1      | 4.3 ± 0.5  |
| Horseradish (1ATJ)         | 41          | 4 ± 1      | 4.3 ± 0.6  |
| Soybean (1FHF)             | 43          | 4 ± 1      | 4.4 ± 0.5  |
| Peanut (1SCH)              | 42          | 3.8 ± 0.9  | 4.0 ± 0.5  |

References

1. Passardi F, Longet D, Penel C, Dunand C. The class III peroxidase multigenic family in rice and its evolution in land plants. Phytochemistry 2004; 65:1879-93; PMID:15279994; http://dx.doi.org/10.1016/j.phytochem.2004.06.023
2. Almagro L, GómezRos LV, Belchi-Navarro S, Bru R, BarcelóRos A, Pedroño MA. Class III peroxidases in plant defense reactions. J Exp Bot 2009; 60:377-90; PMID:19073963; http://dx.doi.org/10.1093/jxb/ern277
3. Tognolli M, Penel C, Greppin H, Simon P. Analysis and expression of the class III peroxidase large gene family in Arabidopsis thaliana. Gene 2002; 288:129-38; PMID:12034502; http://dx.doi.org/10.1016/S0378-1119(02)00465-1
4. Veitch NC. Horseradish peroxidase: a modern view of a classic enzyme. Phytochemistry 2004; 65:249-59; PMID:14751298; http://dx.doi.org/10.1016/j.phytochem.2003.10.022
5. Wright H, Nicell JA. Characterization of soybean peroxidase for the treatment of aqueous phenols. Bioresour Technol 1999; 70:69-79; http://dx.doi.org/10.1016/S0960-8524(99)00007-3
6. Henriksen A, Mirza O, Indiani C, Telum K, Smulevich G, Welinder K, et al. Structure of soybean seed coat peroxidase: a plant peroxidase with unusual stability and haem-apoprotein interactions. Protein Sci 2001; 10:108-15; PMID:11266599; http://dx.doi.org/10.1110/ps.37301
7. Kelley LA, Stenberg MJE. Protein structure prediction on the web: a case study using the Phyre server. Nat Protoc 2009; 4:363-71; PMID:19247286; http://dx.doi.org/10.1038/nprot.2009.2
8. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 2008; 9:40-8; PMID:18215316; http://dx.doi.org/10.1186/1471-2105-9-40
9. Krissinel E, Hendrick K. Secondary-structure matching (SSM): a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D Biol Crystallogr 2004; 60:2256-68; PMID:15572779; http://dx.doi.org/10.1107/S0907444904026460
10. Kabash W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983; 22:2577-637; PMID:6667333; http://dx.doi.org/10.1002/bip.3602211
11. Henriksen A, Welinder KG, Gajhede M. Structure of barley grain peroxidase refined at 1.9 Å resolution. J Biol Chem 1998; 273:2241-8; PMID:9442087; http://dx.doi.org/10.1074/jbc.273.4.2241
12. Veitch NC. Structural determinants of plant peroxidase function. Phytochem Rev 2004; 3:3-18; http://dx.doi.org/10.1023/B:PHYT.0000047799.17604.94
13. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. J Comput Chem 2005; 26:1781-802; PMID:16222654; http://dx.doi.org/10.1002/jcc.20289

14. Lin HH, Tseng LY. DBCP: a web server for disulfide bonding connectivity pattern prediction without the prior knowledge of the bonding state of cysteines. Nucleic Acids Res 2010; 38:W503-7; PMID:20530534; http://dx.doi.org/10.1093/nar/gkq514

15. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 2004; 60:2126-32; PMID:15572765; http://dx.doi.org/10.1107/S0907444904019158

16. Laberge M, Huang Q, Schweitzer-Stenner R, Fidy J. The endogeneous calcium ions of HRP C are required to maintain the functional nonplanarity of heme. Biophys J 2003; 84:2542-52; PMID:12668462; http://dx.doi.org/10.1016/S0006-3495(03)75059-0

17. Cardinali A, Tursi N, Ligorio A, Giuffrida MG, Napolitano L, Calandrino R, et al. Purification, biochemical characterization and cloning of a new cationic peroxidase isoenzyme from artichoke. Plant Physiol Biochem 2011; 49:395-403; PMID:21345687; http://dx.doi.org/10.1016/j.plaphy.2011.01.028

18. Gordon JC, Myers JB, Folta T, Shoja V, Heath LS, Onufriev A. H+ : a server for estimating pKas and adding missing hydrogens to macromolecules. Nucleic Acids Res 2005; 33:W368-71; PMID:15980491; http://dx.doi.org/10.1093/nar/gki464

19. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck J, et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. J Phys Chem B 1998; 102:3586-616; http://dx.doi.org/10.1021/jp973084f

20. Beglov D, Roux B. Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations. J Chem Phys 1994; 100:9050-63; http://dx.doi.org/10.1063/1.466711

21. Marchand S, Roux B. Molecular dynamics study of calbindin D9k in the apo and singly and doubly calcium-loaded states. Proteins 1998; 33:265-84; PMID: 9779793; http://dx.doi.org/10.1002/(SICI)1050-9219(19981101)33:3<265::AID-PROT13.0.CO;2-I

22. Jorgensen W, Chandrasekhar J, Madura J, Impey R, Klein M. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983; 79: 926-35; http://dx.doi.org/10.1063/1.445869

23. Darden T, York D, Pedersen L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. J Chem Phys 1993; 98:10089-92; http://dx.doi.org/10.1063/1.464397

24. Nosé S, Klein M. Constant pressure molecular dynamics for molecular systems. Mol Phys 1983; 50:1055-76; http://dx.doi.org/10.1080/00268978300102851

25. Andersen H. Molecular dynamics simulations at constant pressure and/or temperature. J Chem Phys 1980; 72:2384-93; http://dx.doi.org/10.1063/1.439486

26. Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. J Appl Phys 1981; 52:7182-90; http://dx.doi.org/10.1063/1.328693

27. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph 1996; 14:33-8; PMID:8744570; http://dx.doi.org/10.1016/0267-8855(96)00018-5