Molecular determinants for selective $C_{25}$-hydroxylation of vitamins $D_2$ and $D_3$ by fungal peroxygenases

Fátima Lucas,¹,a Esteban D. Babot,¹,b Marina Cañellas,¹,a,c José C. del Rio,¹ Lisbeth Kalum,² René Ullrich,³ Martin Hofrichter,⁴ Victor Guallar,⁵ Angel T. Martínez⁶ and Ana Gutiérrez⁷

Hydroxylation of vitamin D by Agrocybe aegerita and Coprinopsis cinerea peroxygenases was investigated in a combined experimental and computational study. 25-Monohydroxylated vitamin $D_3$ (cholecalciferol) and $D_2$ (ergocalciferol), compounds of high interest in human health and animal feeding, can be obtained through reaction with both fungal enzymes. Differences in conversion rates and regioselectivity were nevertheless observed, and, to rationalize the results, diffusion of $D_2$ and $D_3$ on the molecular structure of the two enzymes was performed with PELE software. In good agreement with experimental conversion yields, simulations indicate more favorable energy profiles for the substrates’ entrance in $C. cinerea$ than for $A. aegerita$ enzyme. Furthermore, GC-MS analyses show that while a full regioselective conversion into the active $C_{25}$ form is catalyzed by $C. cinerea$ peroxygenase for $D_2$ and $D_3$, $A. aegerita$ yielded a mixture of the hydroxylated $D_3$ products. From the molecular simulations, relative distance distributions between the haem compound I oxygen and $H_{24}/H_{25}$ atoms (hydrogens on $C_{24}$ and $C_{25}$, respectively) were plotted. Results show large populations for O-$H_{25}$ distances below 3 Å for $D_2$ and $D_3$ in $C. cinerea$ in accord with the high reactivity observed for this enzyme. In $A. aegerita$, however, cholecalciferol has similar populations (below 3 Å) for O-$H_{23}$ and O-$H_{24}$ which can justify the small degree of hydroxylation observed in $C_{24}$. In the case of ergocalciferol, due to the bulky methyl group in position $C_{24}$, very few structures are found with O-$H_{24}$ distances below 3 Å and thus, as expected, reaction was not observed in this position.

Introduction

Selective oxygenations of aliphatic compounds are among the most challenging reactions in organic chemistry for the regio and/or stereo specific synthesis of pharmaceuticals and fine chemicals. Monoxygenases catalyzing such hydroxylation reactions include cytochromes P450, a family of haem proteins playing a variety of physiological roles but often requiring an auxiliary flavoenzyme (or flavin-containing module) and a source of reducing power to be activated by $O_2$; two facts that limit their biotechnological applicability. Recently, a new peroxidase type, which shares the active-site architecture and reaction mechanism of cytochromes P450, but has the advantage of being activated directly by $H_2O_2$, was isolated from Agrocybe aegerita.¹² and later identified in a variety of sequenced basidiomycete genomes including that of Coprinopsis cinerea.¹¹ Due to the above characteristics, these unspecific peroxygenases (EC 1.11.2.1) have a huge biotechnological potential as self-sufficient monoxygenases,¹³ for hydroxylation of both aromatic¹⁴,₂₄,₂₅,₂₉,₃₂,₃₉,₄₁ and aliphatic compounds.¹₃,₃₁

The $A. aegerita$ enzyme has been the most widely investigated basidiomycete peroxygenase, but recent studies have shown that the $C. cinerea$ enzyme has comparative advantages related to its high conversion yield/selectivity for some hydroxylation reactions, and its production as a recombinant protein in an industrial expression host (by Novozymes, Bagsvaerd, Denmark).¹₄,₂₅ Hydroxylation of vitamin D for the selective production of its active $C_{25}$-hydroxylated derivatives is one of the reactions where the $C. cinerea$ peroxygenase can be of biotechnological interest (Fig. 1).³ Supplementation with 25-hydroxyvitamin D has a positive effect in different human diseases,³⁹,₄₀ and also raises considerable interest for feeding broiler chickens¹₂,¹₅,₂₁,₃₀ and other farm animals,³⁷ to reduce skeleton problems caused by rapid growth and reduced mobility. Therefore, the use of a peroxygenase in vitamin D hydroxylation represents an attractive alternative to the chemical synthesis.

Fig. 1 Enzymatic conversion of cholecalciferol (vitamin $D_3$: 1) into its active form 25-hydroxycholecalciferol (2) (structures in black) by peroxygenase. Depicted in red are the differences of ergocalciferol’s structure relative to cholecalciferol.

The goal of the present work was to rationalize the differences observed in cholecalciferol and ergocalciferol (vitamins $D_2$ and $D_3$, respectively) conversion rates and regioselectivity by the $A. aegerita$ and $C. cinerea$ peroxygenases. The work presented here consists first of
the enzymatic conversion of these compounds, under optimized conditions, and gas chromatography-mass spectrometry (GC-MS) analyses to identify all reaction products. Then, energy profiles and binding modes of vitamins D$_2$ and D$_3$ were determined by structure-based computational simulations, using the PELE software. Finally, differences in regioselectivity were investigated through the analysis of the most favorable binding orientations in the active site. Results show that molecular simulations can effectively discriminate experimentally observed differences in conversion rates and regioselectivity.

**Results and discussion**

**Experimental hydroxylation reactions**

Conversion of cholecalciferol and ergocalciferol was experimentally determined for both the *A. aegerita* and *C. cinerea* peroxygenases. In the case of cholecalciferol, GC-MS analyses of the reaction mixture revealed that this compound was completely (100%) converted by the *C. cinerea* enzyme within 60 min reaction (Fig. 2C) as compared with the control reaction without peroxygenase (Fig. 2A). However, in the *A. aegerita* peroxygenase reaction only 67% conversion was observed (Fig. 2B). In the case of ergocalciferol (Figs. 2D-F), a conversion of 89% was produced by *C. cinerea* while in *A. aegerita* reaction only 46% product was observed.

Moreover, cholecalciferol and ergocalciferol conversion by the *C. cinerea* peroxygenase showed a strict regioselectivity since it gave exclusively 25-hydroxycalciferol (89 and 100% yield, respectively). Likewise conversion by the *A. aegerita* enzyme of ergocalciferol yielded exclusively 25-hydroxyergocalciferol but for cholecalciferol a mixture of products was observed. The products include 25-hydroxycholecalciferol (50% of the initial substrate) together with 24-hydroxycholecalciferol (8%) and 26/27-hydroxycholecalciferol (9%), as confirmed by comparison with true standards showing identical retention times and mass spectra (Table 1). The double peaks observed in the chromatograms for both substrate and product (Fig. 2) correspond to the isopyro (19β, 9β) and pyro (19α, 9α) isomers formed by thermal rearrangement involving ring-B closure, that vitamin D and its hydroxylated derivatives undergo due to the temperature at which GC-MS EI(+) is carried out. Indeed, the presence of the two isomers during GC separation is a useful indication that a secosteroid of the vitamin D type was injected into the system. The position of the hydroxyl group at the target C$_{25}$ position was established by MS of the TMS derivative. The spectrum shows a prominent ion from C$_{25}$-C$_{35}$ bond cleavage, with characteristic fragment at m/z 131 and molecular ion at m/z 544. Additionally, characteristic fragments at m/z 529 ([M-15]$,^+$), m/z 454 ([M-90]$^+$), m/z 439 ([M-90-151]$^+$), m/z 349 ([M-90-90-151]$^+$) and m/z 413, were also present. Therefore, both the chromatographic profiles and the mass spectra correspond to the isomeric secosteroids.

![Fig. 2](image) GC-MS analyses of cholecalciferol (left) and ergocalciferol (right) hydroxylation by the *A. aegerita* (B, E) and *C. cinerea* peroxygenases (C, F), compared with a control without enzyme showing the substrate peaks (A, D), as TMS derivatives from 60-min reactions. In all cases the isopyro (left) and pyro (right) isomers from secosteroid thermal rearrangement are obtained (two small peaks in A and D correspond to minor additional isomers).

<table>
<thead>
<tr>
<th></th>
<th><em>A. aegerita</em> peroxygenase</th>
<th><em>C. cinerea</em> peroxygenase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Remaining substrate</td>
<td>33.4</td>
<td>0</td>
</tr>
<tr>
<td>24-Hydroxycholecalciferol</td>
<td>8.2</td>
<td>0</td>
</tr>
<tr>
<td>25-Hydroxycholecalciferol</td>
<td>49.8</td>
<td>100.0$^1$</td>
</tr>
<tr>
<td>26/27-Hydroxycholecalciferol</td>
<td>8.6</td>
<td>0</td>
</tr>
</tbody>
</table>

$^1$Over 90% after only 30 min reaction.
**Peroxygenase structure**

A superimposition of the general structure and the haem pocket residues in the *A. aegerita* and *C. cinerea* peroxygenases is shown in **Fig. 3**. The main differences are two longer loops (**Fig. 3A**, arrows) and the substitution of Phe69 by Met69 (**Fig. 3B**, red label) in the *C. cinerea* enzyme. All other haem pocket residues, including the proximal cysteine acting as the fifth ligand of the haem iron (Cys36) and the distal glutamic acid and arginine involved in haem activation by H$_2$O$_2$ (Glu169 and Arg189),$^{32}$ are conserved in the two enzymes. Moreover, differences in the channel providing access to the haem cofactor and the neighbor residues at the channel entrance are shown in **Fig. 4**.

![Fig. 3](image-url) **Fig. 3** Comparison of peroxygenase molecular structures. **A** General superimposition with the *A. aegerita* and *C. cinerea* proteins as pink and yellow ribbons, respectively. The haem group (CPK-colored spheres) and several haem pocket residues (pink- and CPK-colored sticks, respectively) are shown, and two larger loops in the second enzyme are indicated by arrows. **B** Haem pocket residues in the *A. aegerita* (pink sticks) and *C. cinerea* (CPK-colored sticks) peroxygenases (red label indicates non-conserved methionine in position 69 of the latter enzyme). A hypothetical product molecule identified as 4-hydroxymethylimidazole is shown in **A** (brown-colored vdW spheres) and **B** (MZO, brown-colored sticks).$^{32}$ From *A. aegerita* PDB 2YOR, and *C. cinerea* homology model (provided as Supplemental file 1).

**Computational modeling**

**Ligand diffusion energy profiles**

For the simulations of cholecalciferol and ergocalciferol access to the H$_2$O$_2$-activated haem in the *A. aegerita* and *C. cinerea* peroxygenases, the haem cofactor was modeled as compound I. The substrate was placed close to the entrance of the haem-access channel of the proteins prepared at the optimal pH for peroxygenase activity (pH 7). This initial location was identified using SiteMap,$^{36}$ and from there, the ligand was spawned inside the protein by PELE,$^8$ following the distance between the reactive O atom in the haem compound I and the cholecalciferol/ergocalciferol C$_{25}$ atom (**Fig. 1**).

![Fig. 4](image-url) **Fig. 4** Solvent access surfaces showing differences in the size of the haem (CPK-colored sticks) access channel in the peroxygenases of *A. aegerita* (**A**) and *C. cinerea* (**B**). Several neighbor residues are shown as CPK-colored vdW spheres, including Gly323 and Ser324 contributing to occlude the haem channel in the *C. cinerea* enzyme.
PELE simulations were done in two stages: first the substrate is perturbed to reduce the C25-O distance and, when this distance is below 5 Å, the substrate is free to explore the active site cavity (at the haem distal side) with a 15 Å restrain. In the first step, the ligand is perturbed with a combination of large and small translations and rotations, ranging from 0.5 to 1.5 Å for translations, and 0.05 to 0.25 radians for rotations. However, during the second stage, translation range is reduced (0.75-0.25 Å) to perform a finer active site exploration. The plots shown in Fig. 5 correspond to three 48 h simulations each with 80 processors, and show the substrate-C25 to haem-O distance vs. the interaction energy between the protein and the substrate at each of the different poses explored.

![Interaction energies vs. ligand distances from PELE simulations for cholecalciferol (left) and ergocalciferol (right) entrance by the C25 end in the peroxygenases from A. aegerita (A, B) and C. cinerea (C, D). The distances shown (Å) are between the reactive O atom in the haem compound I and the calciferol C25 (for substrate numbering see Fig. 1).](image)

Simulations show that for the A. aegerita peroxygenase the entrance of the substrates from the surface of the protein is quite favorable but then, the access to the activated haem is obtained against an uphill potential (Figs. 5A, B). For the C. cinerea enzyme the entrance is less open than in the A. aegerita peroxygenase and, for this reason, we observed a constrained access to the protein (at around 12 Å). However, the overall energy profile is more favorable, in particular for cholecalciferol. (Fig. 5C). Once inside the protein, the ligand must surpass smaller barriers to reach the haem. From Fig. 5 it is clear that C. cinerea peroxygenase has the most favorable energy profiles and well defined minima in the active site (with C25-O distance around 3 Å). Fig. 5A/5C show that binding of cholecalciferol at the haem site, is more favorable in the C. cinerea peroxygenase (-65 kcal/mol) than in the A. aegerita enzyme (-40 kcal/mol). This difference comes from the fact that the first protein has a tighter binding pocket (see below). In the case of ergocalciferol Fig. 5B/5D, also binding is more favorable in the C. cinerea (-42 kcal/mol) than in the A. aegerita enzyme (-30 kcal/mol). These differences in the energy profiles, which indicate a more favorable protein-ligand interaction in C. cinerea, can explain the higher conversion rate observed for the two compounds in this peroxygenase.

**Ligand binding**

If we overlap cholecalciferol’s positions for the entire PELE simulations for both proteins, we find two main orientations in the active site, as shown in Fig. 6. In Fig. 7, the same analysis for ergocalciferol can be found.

![Superposition of the ligand’s (cholecalciferol) main active site positions (as liquorice) obtained in PELE simulations on the A. aegerita (A) and C. cinerea (B) peroxygenases. Haem cofactor also in liquorice and selected residues as vDW spheres. The protein’s active site access is shown as a surface.](image)

The two minima observed in cholecalciferol’s diffusion in the C. cinerea peroxygenase (Fig. 5C) can also be seen in Fig. 6B. One of the minima (in green) is found at binding distance to the haem O, and a second (in blue), about 5 Å away from the haem oxygen (Fig. 5C). The ligand’s positions observed in the C. cinerea peroxygenase are also found in the A. aegerita enzyme (Fig. 6A), although they do not correspond to energy minima (Fig. 5A). Likewise, ergocalciferol can adopt two positions in the binding pocket (Fig. 7).
When cholecalciferol is in an optimal reacting position, it is held in place by Val77, Phe121, Ala195, Phe199 and Glu196 in the C. cinerea peroxygenase (Fig. 8B), and by Phe69, Ala77, Phe121, Thr192, Gly195, Phe199 and Glu196 in the A. aegerita enzyme (Fig. 8A). Noteworthy that the C. cinerea peroxygenase Met69, homologous to A. aegerita Phe69, is placed away from cholecalciferol and does not appear to have any effect on its position in the active site. Ergocalciferol, however, with an extra methyl group in position C24 and with a C22-C23 double bond, is positioned in a slightly different manner in the active site (Fig. 9), and residues in position 69 (Phe and Met) now interacts with the ligand (red arrows in Fig. 9). This extra constraint could be responsible for the lower reactivity observed for D2 in both peroxygenases.

Inspection of the haem entrance when both ligands are in the active site, reveals a better wrapping of the protein around the ligands in C. cinerea peroxygenase, compared with the A. aegerita enzyme. This in mainly due to the larger loop, where Gly323 is located, which is smaller in the A. aegerita peroxygenase (Fig. 3A, black arrow). These are better illustrated in the haem-access channel of both peroxygenases in Fig. 4, where the position of the above Gly323 is shown. The narrower access to the haem in C. cinerea is also the result of several hydrophobic amino acid substitutions. Replacements to larger side chains in C. cinerea are dominant with: Ala73/Ile73, Ala77/Val77, Gly195/Ala195, Gly241/Ala246 and Val244/Leu249; and only one to a smaller amino acid, Ser240/Gly245. Due to the large entrance cavity of the A. aegerita peroxygenase, both substrates remained solvent exposed at the C2 end. In contrast, more favorable interactions are established at the final substrate position inside the tighter channel of the C. cinerea peroxygenase, which presents extra interactions on the protein surface. The combination of the surface interactions, along with a tighter haem cavity, result in the improved interaction energies seen for D2 and D3 in this protein.
Regioselectivity
To investigate the different regioselectivities observed for D$_2$ and D$_3$ hydroxylation by *A. aegerita* and *C. cinerea* peroxygenases, we have analysed the relative distance distribution of the substrates’ reactive hydrogen atoms in the active site. We have considered as reactive those that can approach the haem compound I oxygen atom close enough to react. Thus, we have taken into account hydrogen atoms in positions C$_{24}$, C$_{25}$, C$_{26}$ and C$_{27}$ for both ligands, in addition to C$_{28}$ hydrogen atoms for ergocalciferol. We have selected all structures (from the PELE simulations) where the distance between H-O is below 5 Å and interaction energies below -20 kcal/mol for D$_3$ and -30 kcal/mol for D$_2$. When the relative frequency of these distances was computed (Fig. S1 and additional results included in SI) it can be seen that for *C. cinerea* the O-H$_{24}$ frequency is dominant for both compounds. The percentage of O-H$_{24}$ distances below 3 Å in *C. cinerea* is 54.5% for D$_3$ and 36.2% for D$_2$, whereas in *A. aegerita* it is 27.4% and 25.7%, respectively. Moreover, in the D$_3$ simulations in *A. aegerita*, the fraction of structures with O-H$_{23}$ below 3 Å is 19.3% which is quite high when compared to the other cases. In fact, the fraction of O-H$_{23}$ distances is only 1.4 times superior to O-H$_{24}$ while for the other systems it ranges between 10 to 20 times. This higher fraction of reactive O-H$_{23}$ distances can explain the observed formation of C$_{25}$ hydroxylated products in cholecalciferol instead of the full regioselective reactions observed in the remaining systems.

Conclusions
Atomic level simulations have been used here to rationalize the differences observed for ergocalciferol and cholecalciferol’s conversion rates and regioselectivity in two peroxygenases. The overall improved energy profiles in *C. cinerea*, the presence of favorable minima, and a high fraction of favorable O-H$_{24}$ distances agrees well with experimentally higher conversion rates.

The main structural differences between the *A. aegerita* and *C. cinerea* peroxygenases that modify the access of cholecalciferol to the activated haem are Ala73/Ile73, Ala77/Val77, Gly195/Ala195, Gly241/Ala246 and Val244/Leu249 changes, all larger amino acids with only Ser240/Gly245 to a smaller one. These larger hydrophobic side chains augment the interaction of D$_2$ and D$_3$ substrate with *C. cinerea*. The better conversion rates observed for *C. cinerea* peroxygenase do not originate in the active site itself (where Phe69 is replaced by Met69), but instead in the ligand access to the active site, and especially in the entrance to the protein. In particular, the larger loop hosting Gly323 creates a barrier that reduces the size of the entrance channel in the *C. cinerea* peroxygenase. This, along with a tighter cavity, is reflected in the more favorable interaction energies. Although the improved reactivity of *C. cinerea* does not appear to be affected by the larger Phe69 side chain, it does hinder the entrance of ergocalciferol relative to cholecalciferol. The presence of an extra methyl group that interacts directly with position 69 in both proteins lowers the reactivity of this compound. Finally, computed relative frequency of distances between the activated haem oxygen and the hydrogen atoms in C$_{24}$ and C$_{25}$ show that, despite the more favourable hydroxylation on tertiary carbons, reaction at cholecalciferol C$_{25}$ occurs when the ratio of favourable O-H$_{23}$/O-H$_{24}$ distances decreases.

Materials and methods
Enzymes and chemicals
Peroxygenase (isoform II) was isolated from *A. aegerita* DSM 22459 grown in soybean medium using a combination of SP-Sepharose chromatography and Mono-P chromatofocusing. *A. aegerita* DSM 22459 is deposited at the [Deutsche Stammsammlung für Mikroorganismen und Zellkulturen](http://www.dsmz.de) Braunschweig (Germany). *C. cinerea* peroxygenase was provided by [Novozymes](http://www.novozymes.com) A/S ( Bagsvaerd, Denmark). The enzyme corresponds to the protein model 7249 from the sequenced *C. cinerea* genome available at the JGI ([http://genome.jgi.doe.gov/Copci1](http://genome.jgi.doe.gov/Copci1)), which was expressed in *Aspergillus oryzae* and purified using a combination of S-Sepharose and SP-Sepharose ion-exchange chromatography (patent WO/2008/119780). One peroxygenase unit is defined as the amount of enzyme oxidizing 1
μmol of veratryl alcohol to veratraldehyde (ε₃₄₀ 9300 M⁻¹·cm⁻¹) in 1 min at 24 °C, pH 7, after addition of 0.5 mM H₂O₂ in the C. cinerea peroxygenase reactions and 2.5 mM H₂O₂ in those with the A. aegerita peroxygenase.

Vitamin D₃, also known as calcidiol or cholecalciferol ([(5Z,7E)-(3S)-9,10-seco-5,7,10(19)-cholestatrien-3-ol; Fig. 1), was tested as substrate of the A. aegerita and C. cinerea peroxygenases. 25-Hydroxyvitamin D₂, also known as calcidiol or 25-hydroxycholecalciferol ([(5Z,7E)-(3S)-9,10-seco-5,7,10(19)-cholestatraene-3,25-diol; Fig. 1), was used as standard for gas chromatography-mass spectrometry (GC-MS) analyses. Vitamin D₃, also known as ercalcidiol or ergocalciferol ([(5Z,7E,22E)-(3S)-9,10-seco-5,7,10(19)]22-ergostatetraen-3-ol; Fig. 1), was also tested as substrate of the two peroxygenases. 25-Hydroxyvitamin D₃, also known as ercalcidiol or 25-hydroxyergocalciferol ([(5Z,7E,22E)-(3S)-9,10-seco-5,7,10(19)]22-ergostatetraene-3,25-diol; Fig. 1), was also used as standard for GC-MS analyses. All compounds were from Sigma-Aldrich.

Enzymatic reactions
Reactions of cholecalciferol and ergocalciferol (0.05 mM) with the A. aegerita and C. cinerea peroxygenases (1 U) were performed in 5 mL of 50 mM sodium phosphate, pH 7, at 40 °C for 60 min (or 30 min), in the presence of 2.5 mM H₂O₂ in the A. aegerita peroxygenase reactions and 0.5 mM H₂O₂ in those with the C. cinerea enzyme. The substrates were previously dissolved in acetone, and added to the buffer (the acetone concentration in the reaction was 40%). In control experiments, the substrates were treated under the same conditions (including H₂O₂) but without enzyme. After the enzymatic reactions, products were recovered by liquid-liquid extraction with methyl tert-butyl ether, dried under N₂, and redissolved in chloroform for GC-MS analyses. Bis(trimethylsilyl)trifluoroacetamide (Supelco) in the presence of pyridine was used to prepare trimethylsilyl (TMS) derivatives. An internal standard was added after the enzymatic reactions to determine product yields.

GC-MS analyses
GC-MS analyses were performed with a Shimadzu QP2010 Ultra equipment, using a fused-silica DB-5HT capillary column (30 m x 0.25 mm internal diameter, 0.1 μm film thickness) from J&W Scientific. The oven was heated from 120 °C (1 min) to 300 °C (15 min) at 5 °C·min⁻¹. The injection was performed at 300 °C, the transfer line was kept at 300 °C, and helium was used as carrier gas. Compounds were identified by mass fragmentography, and by comparing their mass spectra with standards, and quantitation was obtained from total-ion peak area, using molar response factors obtained from cholecalciferol, ergocalciferol, 25-hydroxycholecalciferol and 25-hydroxyergocalciferol standards. The two latter compounds were also used as external standards for calculation of product yields.

PELE computational analyses
The starting structures for PELE simulations were the A. aegerita peroxygenase crystal (2YOR) and a homology model for the C. cinerea peroxygenase structure obtained using 2YOR as template. As the optimum pH for peroxygenase activity is 7, the structures were prepared accordingly using Schrodinger’s Protein Preparation Wizard and H++ web server. Histidines were δ-protonated, with the exception of His82 (ε-protonated) and His118 and His251 (double protonated). All acidic residues were deprotonated, except Asp85 that was kept in its protonated state. The ergocalciferol and cholecalciferol molecules were optimized with Jaguar at the DFT/M06 level with the 6-31G** basis and a PBF implicit solvent in order to obtain their electrostatic potential atomic charges. Finally, the haem site was modeled as thiolate-ligated compound I after being fully optimized in the protein environment with quantum mechanics/molecular mechanics (QM/MM) using QSite. The electronic calculations show three unpaired electrons: two located on the oxoiron group and a third on the heme and less than 1% spin contamination.

Once the initial protein structure was prepared and ligands optimized, these were placed manually in identical positions at the entrance of the haem-access channel and PELE simulations were performed. PELE is a Monte Carlo based algorithm that produces new configurations through a sequential ligand and protein perturbation, side chain prediction and minimization steps, freely available at https://pele.bsc.es. New configurations are then filtered with a Metropolis acceptance test, where the energy is described with an all-atom OPLS force field and a surface generalized Born solvent. In this way it is possible to locate and characterize local and global minima structures for the most favorable protein-ligand interactions. PELE has been successfully used in a number of ligand migration studies with both small and large substrates.

Acknowledgements
This work was supported by the INDOX (KBBE-2013-7-613549) and PELE (ERC-2009-Adg 25027) EU projects, and by the BIO2011-26694 and CTQ2013-48287 projects of the Spanish Ministry of Economy and Competitiveness.

Notes and references
1 These three authors equally contributed to this work
2 Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, Jordi Girona 29, E-08034 Barcelona, Spain. E-mail: fatima.lucas@bsc.es
3 Instituto de Recursos Naturales y Agrobiología de Sevilla, CSIC, Reina Mercedes 10, E-41012 Seville, Spain. E-mail: anagu@irnase.csic.es; Fax: +32 954624002; Tel: +32 954624711
4 Anausomics Biotech, Balmes 89, E-08008 Barcelona, Spain
5 Novozymes A/S, Krogshoejvej 36, 2880 Bagsvaerd, Denmark
6 TU Dresden, Department of Bio- and Environmental Sciences, Markt 23, 02763 Zittau, Germany
7 ICREA, Passeig Lluís Companys 23, E-08010 Barcelona, Spain
Electronic Supplementary Information (ESI) available: Molecular model for *C. cinerea* peroxygenase obtained using the *A. aegerita* crystal structure as template. See DOI: 10.1039/b000000x/

33. Schrödinger. QSite 5.7. LCC, New York, 2011.
34. Schrödinger. SiteMap 2.5. LCC, New York, 2011.