Computational characterization of the substrate-binding mode in coproporphyrinogen III oxidase

Pedro J. Silva[§] *and Maria João Ramos*[†*]

[§] REQUIMTE, Faculdade de Ciências da Saúde, Universidade Fernando Pessoa, Rua Carlos da Maia, 296, 4200-150 Porto-Portugal

[†]REQUIMTE, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 687, 4169-007 Porto – Portugal

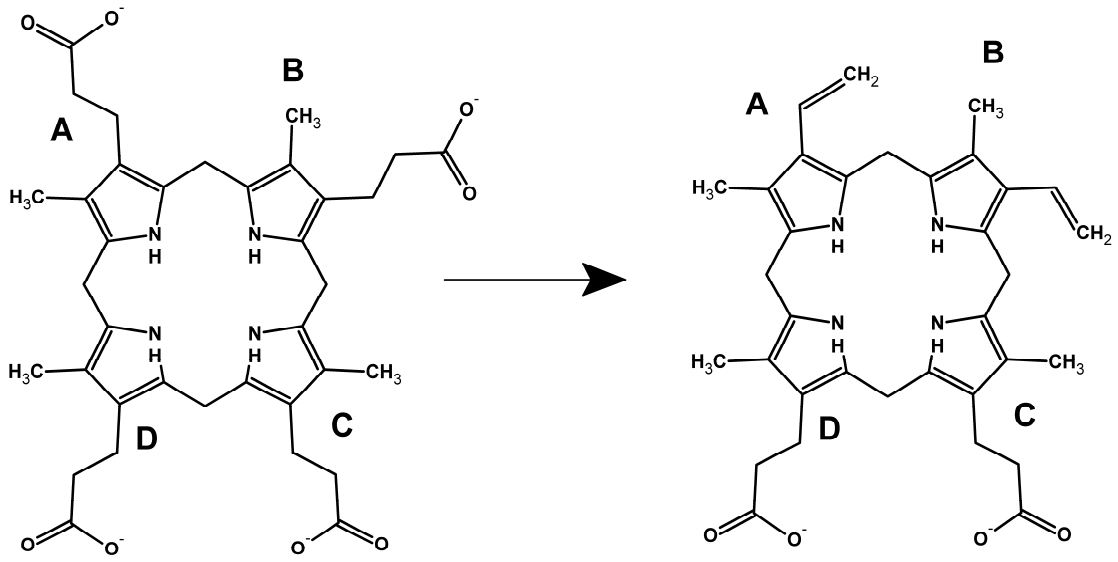
Abstract:

Oxygen-dependent coproporphyrinogen III oxidase catalyzes the sequential decarboxylation of the propionate substituents present on the A and B rings of coproporphyrinogen III in the heme biosynthetic pathway. Although extensive experimental investigation of this enzyme has already afforded many insights on its reaction mechanism, several key features (such as the substrate binding mode, the characterization of the active site and the initial substrate protonation state) remain poorly described. The molecular dynamics simulations described in this paper enabled the determination of a very promising substrate binding mode and the extensive characterization of the enzyme active site. The proposed binding mode is fully consistent with the known selectivity of the active site towards substituted tetrapyrroles, and explains the lack of activity of the H131A, R135A, D274A and R275A mutants and the reasons behind the non-occurrence of catalysis on the C and D rings of the tetrapyrrole. An important role in this binding mode is fulfilled by G276, as its carbonyl oxygen intervenes in the substrate anchoring by hydrogen-bonding its ring D pyrrole NH group. The presence of this interaction (which is only possible with protonated NH pyrrole group), and the absence of positively-charged side chains close to the pyrrole nitrogen (which might stabilize the N-deprotonated pyrrole postulated in some mechanistic proposals) shows that the pyrrole ring is very unlikely to undergo deprotonation during the catalytic cycle and allow the discrimination between the previously postulated mechanistic proposals.

KEYWORDS: molecular dynamics, coproporphyrinogen III oxidase, active site characterization

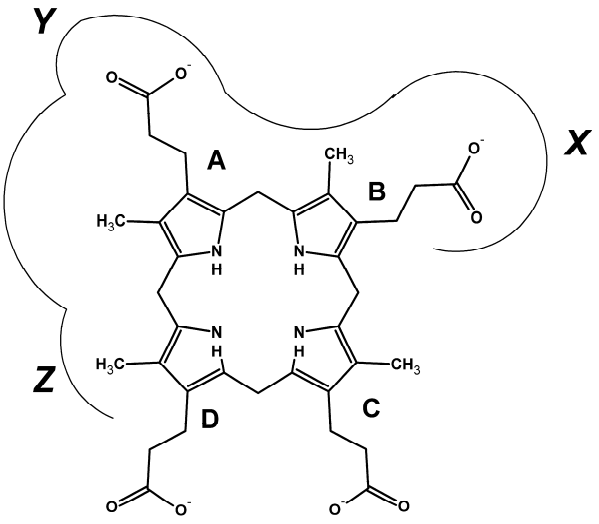
*corresponding author: mjramos@fc.up.pt

I. Introduction



The oxygen-dependent coproporphyrinogen oxidase (EC 1.3.3.3) is a homodimeric enzyme lacking any cofactors[1,2] which catalyzes the sequential decarboxylation of the propionate substituents present on the A and B rings of coproporphyrinogen III[3-7] (Scheme 1). Deficient activity of coproporphyrinogen III oxidase causes the accumulation and non-enzymatic oxidation of coproporphyrinogen, leading to neurological disturbances and skin photosensitivity[8]. The enzyme functions as a dimer in solution, and both the human[9] and yeast[10] enzymes have been crystallized and solved at good resolutions (2.0 Å for the yeast enzyme, and 1.58 Å for the human variant). Site-directed mutagenesis experiments have shown that several conserved residues (H131[11], as well as R135, D274 and R275[12] – amino acids numbered according to the mature yeast protein) are critically important for substrate binding and/or catalysis, and extensive investigations with synthetic pyrroles[5,13-16] have further shown that the active site contains different regions important for substrate recognition: the "Z" region can accommodate small nonpolar groups (R=H, methyl, ethyl vinyl) and discriminates against charged groups (like acetate and propionate), a neighbouring "Y" region binds the propionate group undergoing oxidation, and a third "X" region requires the presence of a propionate (or vinyl[16]) group (Scheme 2). Due to these constraints, only tetrapyrroles bearing a R-Me-Pr-Me-Pr/Vin sequence of substituents (where R is a small nonpolar group) may be metabolized by the enzyme. The first decarboxylation may occur even if the propionate groups present on the unreacting C and D rings are replaced by ethyl groups, but this

substitution prevents the second decarboxylation[5,17], which shows that after the first decarboxylation the substrate either leaves the active site or undergoes a rotation in order to optimize binding affinity. The recognition of the substitutents on the coproporphyrinogen rings C and D also affects the kinetic parameters of the enzyme[16,17], although the reasons for this behaviour have not yet been ascertained.



In a previous report[18] we used quantum-mechanical computations on small models to analyze the kinetic feasibility of several alternative mechanistic proposals proposed in the literature. We found that a one-electron oxidation of the substrate by oxygen, followed by H atom abstraction from the propionate sidechain by the superoxide anion thus formed (the Arigoni model[9]) may only occur if the substrate pyrrole ring is deprotonated, whereas a direct attack of an oxygen molecule by the pyrrole ring (the Lash model[19]) is possible regardless of the protonation state of the reacting pyrrole (Figure 1). The discrimination between these proposals requires knowledge of the pKa of the pyrrole rings in the enzyme-bound substrate and could not be performed at the time due to the unavailability of a three-dimensional structure of the substrate-bound enzyme. In this work we have circumvented this problem by performing a computational substrate docking followed by molecular dynamics simulations and binding site analysis. The results enabled the determination of the substrate binding mode and the extensive characterization of the enzyme active site, and the discrimination between the postulated mechanistic proposals.

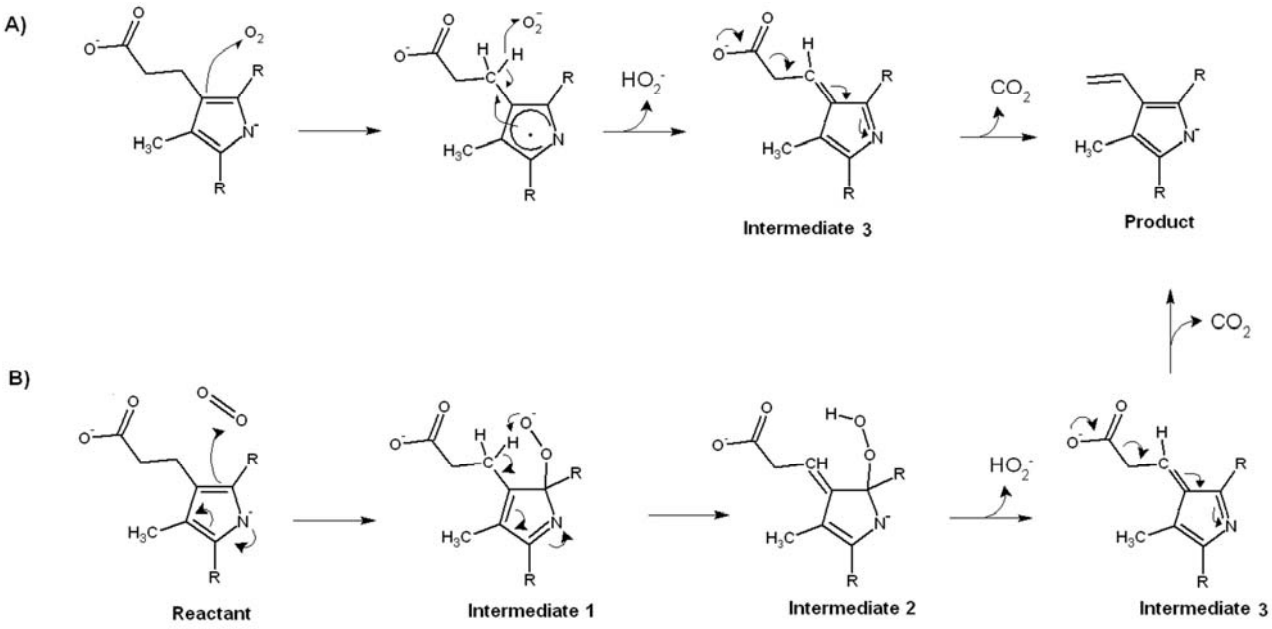


Figure 1: Proposed reaction mechanisms for coproporphyrinogen oxidase. A) Arigoni model[9]. B) Lash model[19]. The Arigoni model is energetically feasible only if the pyrrole ring nitrogen is deprotonated, whereas the Lash model may occur with either protonated or deprotonated substrate, although for simplicity only the mechanism for the deprotonated state is shown.

II. Methods

Molecular docking

All computations were performed with the yeast coproporphyrinogen oxidase dimer structure[10]. The AutoDock program (version 4.0.1; The Scripss Research Institue, la Jolla, CA) was used in preliminary attempts to determine the putative binding site of coproporphyrinogen to coproporphyrinogen oxidase. A grid centered on the coproporphyrinogen oxidase dimer with 127×127×127 points with a spacing of 0.525 Å was used. Each docking experiment (using a genetic algorithm) consisted of an initial population of 250 individuals, 100 runs, and 6.5×10^6 energy evaluations. Due to the high flexibility of

the substrate and to the narrow dimensions of the binding site, docking was performed with an unsubstituted cyclic tetrapyrrole, rather than with the full coproporphyrinogen substrate. The docking poses obtained were then used as starting points for molecular dynamics simulations after adding the appropriate substitutents to each pyrrole ring.

Molecular dynamics

All molecular dynamics simulations were run with YASARA[20] with the AMBER99[21] forcefield under periodic-boundary conditions and with explicit water, using a multiple time step of 1.25 fs for intramolecular and 2.5 fs for intermolecular forces. Simulations were performed in cells 10 Å larger than the protein dimer along each axis (final cell dimensions 116 × 94 × 77 $Å^3$) , and counter-ions were added to a final concentration of 0.9 % NaCl. A 7.86 Å cutoff was taken for Lennard-Jones forces and the direct space portion of the electrostatic forces, which were calculated using the Particle Mesh Ewald method[22] with a grid spacing <1 Å, 4th order B-splines and a tolerance of 10^{-4} for the direct space sum. Simulated annealing minimizations started at 298 K, velocities were scaled down with 0.9 every ten steps for a total time of 5 ps. After annealing, simulations were run at 298 K. Temperature was adjusted using a Berendsen thermostat[23] based on the time-averaged temperature, i.e., to minimize the impact of temperature control, velocities were rescaled only about every 100 simulation steps, whenever the average of the last 100 measured temperatures converged. Substrate parameterization was performed with the AM1BCC protocol[24], which assigns atomic charges by applying simple additive bond charge corrections (BCCs) to AM1 atomic charges and has been shown to provide a fast and robust alternative to HF/6-31G* ESP-fit charges for ligand parameterization for molecular dynamics applications. Protein sidechain protonation patterns were assigned according to Krieger *et al.*[25]. All simulations were run for at least 10 ns after an initial 500 ps equilibration.

III. Results and Discussion

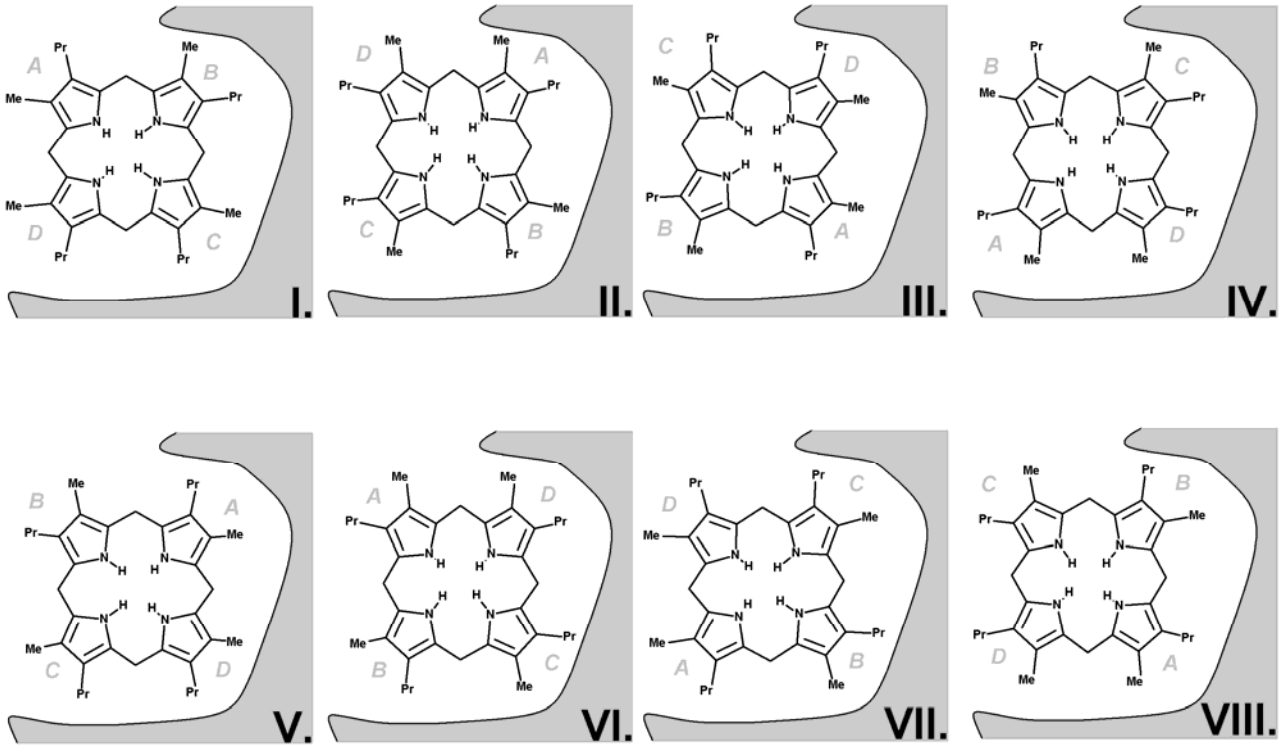


Figure 2: The eight possible substrate orientations in coporporphyrinogen III oxidase active site cavity.

The distribution of the pyrrole substituents around the coproporhyrinogen III molecule is not fully symmetric, as the conformation of the D ring has been flipped earlier in the heme biosynthetic pathway in the reaction catalyzed by uroporphyrinogen III synthase[26,27]. Since the small asymmetry introduced into the molecule consists only in the interchange of a methyl for a propionate group (and vice-versa), the eight possible binding modes to the active site (Figure 2) are expected to show broad similarities and therefore the full characterization of the active site must take into account not only the determination of the appropriate region of the enzyme surface, but also the precise orientation of the assymetrical substrate in regard to the active site. The specificity of the enzyme to tetrapyrroles bearing only certain substitution patterns further demands that the characterization of the possible binding sites be as thorough as possible, and that as much attention be paid to the differences as to the similarities between the different symmetry-related binding poses obtained.

The available crystal structures[9-10] show that coproporphyrinogen III oxidase may exist in two main conformations, distinguishable due to the relative position of helixes H2 (Ser79-Lys87) and H8 (Arg275-Leu281), which may either move apart or towards each other, thereby opening or closing access from the solvent to a cavity which contains the putative active site[10]. Our preliminary docking attempts with AutoDock showed that in the "open" position a tetrapyrrole substrate may fit the solvent-accessible narrow cavity in such a way that the substituents of two of its pyrrole rings remain totally or partially solvent-exposed. However, the cramped dimensions of this putative active site does not allow the free rotation of the substrate during the docking procedure and therefore the search for the correct binding conformation had to be further refined through molecular dynamics simulations. The binding pose obtained from docking experiments with unsubstituted tetrapyrrole was used as starting point for eight different molecular dynamics simulations (after inclusion of pyrrole substituents at the appropriate positions), representing each of the putative substrate orientations. Since the correct binding mode has been experimentally shown to implicate interactions with H131, R135, D274 and R275, additional simulations were performed with an *in silico* mutant where these four amino acids have been changed to alanine. The comparison of the simulations performed in wild-type and mutant proteins assisted our identification of the correct binding mode and improved the characterization of the relevant interactions between the substrate and the protein, as will be shown subsequently.

The dimer structure remained quite stable along the simulations, except in a short unstructured stretch of the N-terminal and in the protein regions involved in the movement of the H2 helix away from/towards helix H8 (Figure 3). (i.e. the opening/closing of the substrate-binding cavity): these include the H2 helix itself (Lys79-Asp89), the loop between this helix and β-sheet S4 (His90-Gly108), and two β-sheet termini contacting this loop: the C-terminus of β-sheet S5 and the N-terminus of β-sheet S6 (Glu138-Trp149). In the substrate-free subunit, the C-terminal of helix H2 remains ≈ 13.1 ± 1.5 Å from the N-terminal of helix H8 along the simulations, whereas its N-terminal lies ≈ 15.5 ± 1.8 Å from the C-terminal of helix H8. In the substrate-bound subunit the average distance between the N-terminal of helix H2 and C-terminal of helix H8 is lower than that in the substrate-free subunit in all

binding modes, whereas the H2 C-terminal to H8 N-terminal distance varies dramatically with the binding mode considered (Table 1) For all binding modes (with the exception of binding modes IV and VI), the quadruple mutation of amino acids H131, R135, R275 and D274 to alanine causes this inter-helical distance to increase dramatically (Table 2), yielding a very open conformation of the active site. As explained in the following section, this is mostly due to movements of the substrate in the active site, rather than in the intrinsic flexibility of the enzyme. In all simulations performed the Root Mean Square Deviations of the structures (after excluding the most flexible region between amino acids 77 and 110) remain low (around 2 Å) (Supporting Information). Most of the remaining variability is due to the unstructured N-terminal and to the loops (45-50 and 138-149) contacting the flexible 77-110 region (See Supporting Information for RMSF graphs).

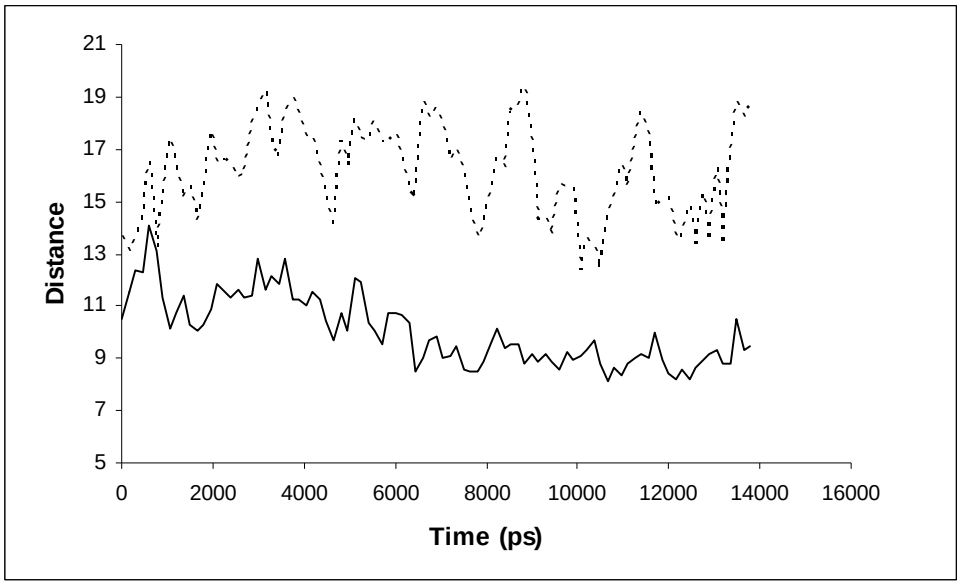


Figure 3: Distances between the termini of helixes H2 and H8 in the substrate-bearing chain of coproporphyrinogen III oxidase (Binding mode VI). Continuous line: Cα 82-Cα 282 (N-terminal H2 – C-terminal H8); Broken line: Cα 89-Cα 275 (C-terminal H2 – N-terminal H8). Additional graphs for all other binding modes (both in wild-type and mutant protein) are available as Supporting Information.

Table 1: Distances (in Å) between the termini of helixes H2 and H8 of substrate-bearing (chain A) and substrate-free (chain B) wild-type coproporphyrinogen III oxidase.

Binding mode	Chain A		chain B	
	Cα 82-Cα 282	Cα 89-Cα 275	Cα 82-Cα 282	Cα 89-Cα 275
I	16.35±1.58	12.60±1.03	14.62±1.36	13.17±1.40
II	12.32±1.39	10.97±0.55	13.69±1.09	13.75±1.13
III	15.93±1.06	11.95±1.17	15.75±1.40	12.88±1.39
IV	8.77±0.78	10.02±0.37	17.05±1.54	12.37±1.20
V	14.53±0.96	11.27±0.66	15.63±1.40	14.25±1.61
VI	9.98±1.31	10.10±0.43	16.21±1.76	12.43±1.22
VII	16.23±1.00	12.32±0.62	16.22±1.64	12.99±0.85
VIII	13.12±2.12	11.08±1.15	15.46±1.45	13.06±1.59
apoprotein	14.33±1.47	12.10±1.09	16.55±1.61	12.40±0.40

Table 2: Distances (in Å) between the termini of helixes H2 and H8 of substrate-bearing (chain A) and substrate-free (chain B) mutant coproporphyrinogen III oxidase.

Binding mode	chain A		chain B	
	Cα 82-Cα 282	Cα 89-Cα 275	Cα 82-Cα 282	Cα 89-Cα 275
I	20.17±1.28	14.16±0.84	15.22±1.91	14.90±0.91
II	13.09±1.41	11.48±0.74	14.11±1.24	14.26±0.80
III	17.45±1.07	13.27±1.05	17.40±1.61	12.83±0.94
IV	11.88±1.02	10.77±0.68	17.54±1.15	12.86±0.80
V	16.23±2.09	13.12±1.27	15.15±1.47	14.87±0.94
VI	10.32±0.79	11.08±0.86	17.04±1.65	12.02±0.93
VII	18.80±1.56	13.46±1.31	17.68±1.53	13.20±1.11
VIII	18.12±1.45	15.02±1.39	16.67±1.54	14.53±1.07

The precise shape and limits of the active site cavity vary with substrate orientation according to the interactions between the pyrrole substituents and the amino acid sidechains: a detailed description of each possible binding mode is therefore provided in the next section. The overall shape of the active site cavity as well as important aminoacids interacting with the substrate in at least one orientation are depicted in Figure 4. In every case, the narrow confines of the substrate-binding crevice prevent the substrate from assuming the "domed" conformation postulated previously[12,19], and impose a chair/chaise longue conformation instead. Unless otherwise noted, the distances mentioned in these descriptions are measured between the terminal carbon of methyl (or propionate) pyrrole substituents and Cγ (Asp residues), Cε (Arg/His residues), sidechain O (Thr/Ser residues), sidechain N (Asn residues) or closest carbon atom (all other residues) and averaged from simulation snapshots taken at 7.5 ps intervals from 500-10500 ps. Where relevant, standard deviations of the distances are also provided.

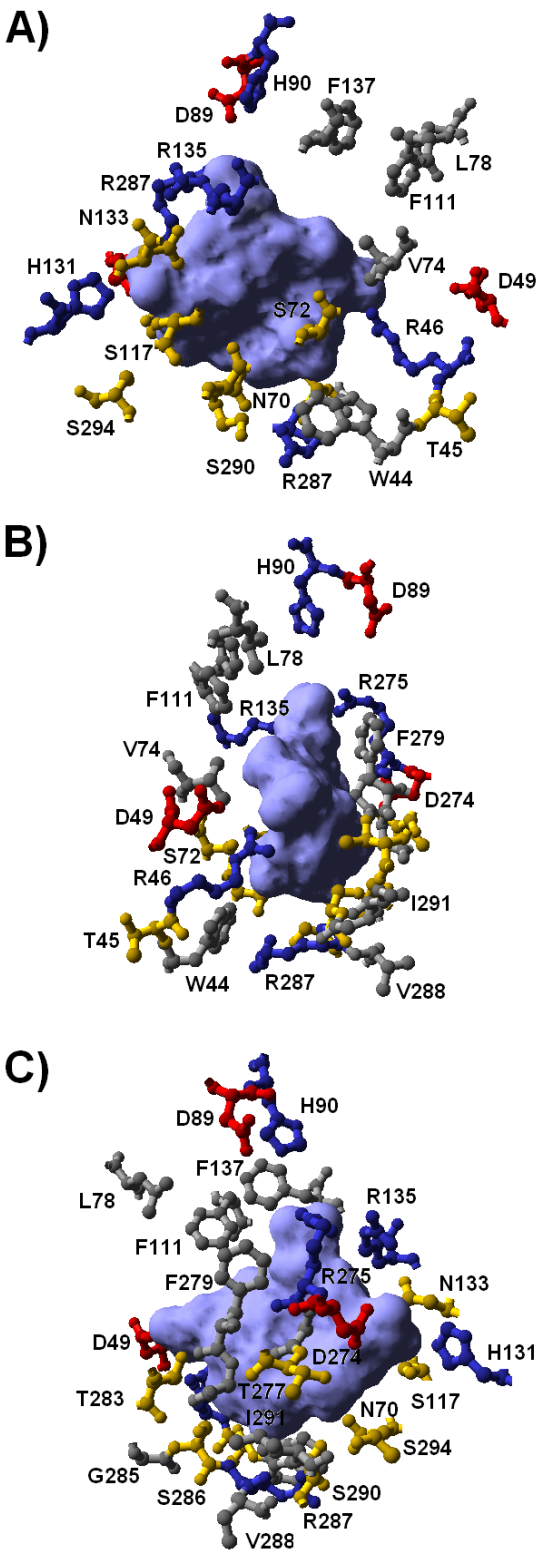


Figure 4: Active site cavity in coproporphyrinogen III oxidase and surrounding amino acids (Positively charged: blue; negatively-charged: red, polar, uncharged: yellow; non-polar: grey). Panels A, B and C are related to each other through sequential ≈90º rotations around the vertical axis

Table 3: Comparison of predicted features of each binding mode with experimentally derived information.

Binding mode	Binding involves H131, R135, D274 and/or R275	Binding changed upon mutation of H131, R135, D274, R275	Compatible with experimental constraints on the substrate substituents
I	No	No	No
II	R135 only	No	No
III	H131 only	Yes	No
IV	R135 only	Yes	No
V	No	Yes	No
VI	R135 only	No	No
VII	No	Yes	No
VIII	H131, R135, R275	Yes	Yes

III.A Binding mode I

This putative binding mode is not energetically favored, as the substrate molecule spontaneously rotates in order to avoid the solvent exposure of the two adjacent hydrophobic methyl groups present on rings A and D. In the stable binding pose obtained upon this rotation the binding site is wider than in other binding modes due to the larger amplitude of the movement of the“lid” formed by the H2 helix and the H2-S4 loop, and the substrate becomes tilted relative to the conformation depicted in Figure 4. The ring A methyl group is now surrounded by Thr283, Gly280 and Phe279. Although both the propionate and the methyl groups present on ring B become solvent-exposed, the methyl group is partially screened (within 4 Å) by Ser72 and Val74. The C ring propionate establishes weaker interactions with Ser117 (4.33 ± 0.56 Å away), Ser294 (4.5 Å), His131 (5.38 ± 0.42 Å) and Asn 133 (5.47 ± 0.45 Å) whereas the A propionate interacts with Arg46 and the D propionate is stabilized

through short-range interactions (<4 Å away) with Ser55, Asn70, Arg287, and Ser290. No changes in binding mode are apparent upon mutation of H131, R135, R275 and D274 to alanine (Supporting information, Figure S1). The solvent-exposure of the ring B methyl group in this binding mode is not consistent with the experimentally verified requirement of a small non-polar substituent (rather than e.g. acetate[13]) at that position. The non-involvement of H131, R135, R275 and D274 in this substrate binding mode also argues against its mechanistical relevance.

III.B Binding mode II

This simulation afforded a very stable binding conformation, with seven different short-range contacts between polar/charged side chains and the substrate propionate groups maintained throughout the experiment: the A propionate remains bound to Arg135 (3.94 ± 0.12 Å) and His90 (4.31 ± 0.25 Å), Ser117 (3.22 ± 0.17 Å) and Asn133 (4.33 ± 0.41 Å) surround propionate B, the C propionate is kept close to Ser72 (3.40 ± 0.16 Å) and Ser 290 (3.79 ± 0.19 Å) though hydrogen bonds, and the exposed D propionate negative charge is stabilized by Arg46 (4.27 ± 0.17 Å). Both the D- and the A-ring methyl groups remain accessible to solvent, though lined with hydrophobic residues (Val74, Phe111 and Phe279 for D methyl, Met86 and Phe137 for A methyl). Apart from a marked increase in the mobility of Arg46, no changes can be seen in the binding mode upon quadruple mutation (Supporting Information, Figure S2).

III.C Binding mode III

In this putative binding mode, the propionate groups appended to the rings C and D remain solvent-exposed throughout the simulation. The entrance to the crevice is now narrower than in the previous binding modes and the protein surface surrounds the D- and A- ring methyl groups. The ring A propionate remains very tightly bound to the active site through strong interactions with Asn133 (3.26 ± 0.14 Å) and His131 (3.49 ± 0.16 Å), whereas the negative charge carried by the ring B propionate is

stabilized by Arg287 (4.02 ± 0.19 Å), Ser72 (4.64 ± 0.52 Å), and Ser290 (4.47 ± 0.77 Å).In the mutated protein the B propionate more easily approaches Ser290 (3.54 ± 0.24 Å), Arg287 and Asn70 due to the removal of the stabilizing His131 that, in the wild type-protein, attracted propionate A and pulled the substrate in the opposite direction (Supporting information, Figure S3). This subtle difference in binding allows the entrance to the crevice to widen, so that in the mutated protein the D-ring is more exposed to solvent and therefore a substituent slightly larger than methyl (e.g. vinyl) may be more easily accommodated at this position than in the WT protein. In the WT protein there is enough free room around the ring B methyl to accommodate polar substituents than methyl (e.g. acetate), nn contrast to the experimental observation[13].

III.DBinding mode IV

In this substrate conformation, only the substituents present on ring B are exposed to the solvent. The substrate remains bound to the enzyme through stable interactions involving all remaining propionates: propionate A interacts with Ser72 (3.30 ± 0.24 Å) and Asn70 (4.143± 0.41 Å), propionate D hydrogen-bonds Asn133 (3.43 ± 0.17 Å) and propionate C is responsible for electrostatic interactions with His90 (3.62 ± 0.15 Å) and Arg135 (4.11 ± 0.15 Å). In the quadruple mutant, the lack of Arg135 allows the substrate to change its binding mode: propionate C is now only attracted by His90 (4.11 ± 0.15 Å), propionate D approaches Asn133 (3.68 ± 0.32 Å), and propionate A slides past Asn70 (which is now left at 7.00 ± 0.46 Å), while remaining close to Ser72 (3.53 ± 0.17 Å) and approaching Arg46 (4.20 ± 0.17 Å) (Supporting Information, Figure S4). The lack of interactions with Arg275 and His131 argues against the relevance of this binding mode to the catalytic effect of the enzyme, as does the solvent exposure of the B methyl group.

III.E Binding mode V

Even in the wild-type protein, binding in this conformation is almost completely due to interactions with only one of the substrate carboxylate groups. In this simulation, propionates A and B remain solvent-exposed throughout, whereas propionate C remains within short range of Ser290 (3.34 ± 0.34 Å) and Arg287 (4.39 ± 0.32 Å). In the first 8 ns of the simulation Asn70 maintains a stable hydrogen-bond with propionate C (≈91% of the time). Although the stability of this bond decreases after this point (to 54% of the time), Asn70 remains quite close to propionate C (average distance in the first 10.5 ns is 3.48 ± 0.28 Å). Propionate D interacts only with Ser117 (3.47 ± 0.39 Å), whereas the A, B and D methyl groups remain quite exposed to the solvent, though surrounded by non-polar residues. In the simulation of the mutated protein, the crevice opens and helixes H2 move away from each other. The substrate partially exits the binding region, and the only surviving binding interactions are those between Arg287 and propionate C (3.83 ± 0.13 Å) and a new (less stable) interaction between this propionate and Ser72 (3.85 ± 1.03 Å). Eventually, a stable interaction between Arg46 and propionate C also develops, but this is not enough to maintain the enzyme-substrate complex tightly bound (Supporting information, Figure S5). The experimental results of enzymatic conversion of substituted porphyrinogens suggest that the active site should have enough free space around the position where the D-ring methyl is accomodated, whereas the methyl group in ring A cannot be replaced by larger substituents[5,13-15]. The relatively open environment around the A methyl group in this binding conformation, the paucity of electrostatic contacts between the protein and substrate, and the lack of interactions involving the experimentally determined amino acids, all argue against the mechanistic relevance of this substrate binding mode.

III.F Binding mode VI

In this binding mode propionate A remains solvent-exposed and bound to Arg46 (4.01 ± 0.16 Å). This stable hydrogen bond is present in 71% of the simulation, and never remains broken for more than 75 consecutive ps. Propionate groups B and C remain close to Asn70 (4.1001 ± 0.45 Å) and Asn133 (3.89

± 0.41 Å), respectively, whereas propionate D remains strongly bound by two stable ionic interactions with His90 (3.70 ± 0.20 Å) and Arg135 (4.13 ± 0.15 Å). The methyl group on ring D is closely surrounded by Leu78, Phe279, Met86 and Phe137, which renders this binding mode (like the previous one) incompatible with the observed catalysis of several D-ring substituted porphyrinogens. In the mutated enzyme the removal of Arg135 releases propionate D, which remains attracted only by His90, the substrate C propionate moves further towards Asn133, and the propionate B approaches Asn70 and Ser72. Propionate A remains anchored to Arg46, and the binding mode maintains high similarity to that observed with the wild-type protein (Supporting Information, Figure S6).

III.G Binding mode VII

Throughout the simulation of this binding mode, the methyl groups from rings A and D remain exposed to the solvent. This behavior is not consistent with the lack of activity of the enzyme on substituted substrates bearing larger substituents than methyl (on ring A) or vinyl (on ring D). The simulation showed this putative binding mode to be the least stable of all those analyzed: only propionate A establishes stable hydrogen bonds with enzyme amino acids (Ser72, 3.43 ± 0.23 Å and Asn70, 4.36 ± 0.41 Å), whereas the interactions between propionate B and Arg135 and Asn 133 do not survive past the initial 20% of the simulation and Arg46 only sporadically interacts with propionate D. In the mutated enzyme, the cavity opens widely and this binding mode becomes even less effective, as the interaction between Asn70 and propionate A disappears, and the distance between Ser72 and propionate A becomes more variable and (on average) longer (4.39 ± 0.61 Å).(Supporting Information, Figure S7).

III.H Binding mode VIII

In the last binding mode, propionates B and C are solvent-exposed, as is the methyl group on the C-ring. The simulation proceeds in a most dramatic way: although the substrate propionate groups begin the simulation quite far from His90, His131, Arg275, several close contacts between these sidechains and the substrate propionates form after only 0.30 ns and create a binding arrangement than remains very stable throughout the simulation. This binding mode agrees the most with the experimental observations on the binding specificity and involved amino acids (Table 3). The substrate becomes very firmly anchored to the binding site through three short, very stable, salt bridges: propionate A-His131 (3.59 ± 0.26 Å) propionate B-Arg275 (4.03 ± 0.27 Å) and propionate C-Arg46 (4.00 ± 0.19 Å). These electrostatic interactions further benefit from very favourable geometric arrangements, which allow hydrogen-bonding in 85% (His131), 90% (Arg 275) and 95% (Arg 46) of the simulation time. Additional interactions between Ser72-propionate D (3.36 ± 0.39 Å) and Asn133-propionate A (4.42 ± 0.50 Å) assist the binding. During ¾ of the simulation, propionate B interacts strongly with His90 (4.63 ± 0.25 Å), but this interaction then weakens (to 5.84 ± 0.59 Å) and is replaced by interactions between this propionate and Arg135 (4.45 ± 0.39 Å in the last nanosecond of the simulation). At the same time, the distance between propionate A and Ser117 also decreases significantly (from 4.98 ± 0.47 Å to 3.42 ± 0.28 Å). In the mutated protein most of these interactions disappear and the substrate remains attached only through ionic/hydrogen-bonding interactions between Arg46 and propionates C and D (4.01 ± 0.31 Å and 4.59 ± 0.36 Å, respectively) and a weak, very labile hydrogen bond between Ser72 and propionate D (4.44 ± 1.00 Å) (Figure 5). In the wild-type protein, an important role is played by the strongly conserved residue G276 in this binding mode: its carbonyl oxygen occupies a central position in the active site, which allows it to establish a stable hydrogen bond with the pyrrole NH groups in ring D (2.07 ± 0.20 Å) and also keep it close to the pyrrole nitrogens in rings C (3.32 ± 0.38 Å) and A (3.88 ± 0.27 Å). This hydrogen bond is quite stable, as it is present in 75% of the simulation snapshots and never remains unbroken for more than 60 consecutive ps. Except for His90, all involved aminoacids described are fully conserved in human, yeast, mouse, puffer fish, fly, *A. thaliana*, *C. reinhardti* and *E. coli*[9]. Analysis of the molecular surface reveals that the environment surrounding the ring D methyl

group in binding pose VIII is not fully occupied by this substituent and that, in agreement with experimental results, this position can accommodate non-polar substituents with two (but not three) carbon atoms.

Asp274 remains close to propionate A throughout the simulation (4.70 ± 0.30 Å away from its carboxylate group, and 4.48 ± 0.30 Å away from its β carbon), but its negative charge is expected to destabilize the substrate carboxylate and its role seems to be unrelated to substrate binding. Two possible roles are suggested by the molecular simulations: on the one hand, its proximity to the propionate may facilitate the decarboxylation reaction (by increasing the energy of the negatively charged reactant state); on the other hand, each carboxylate oxygen in Asp274 establishes a hydrogen bond with a different enzyme amino acid (Tyr196 on the C-terminus of helix H4 and Thr277 on helix H8) and plays therefore a role in the structural stability of the enzyme. Analysis of the Tyr196-Thr277 pair in each of the dimer subunits in the mutant and wild-type simulations (Table 4) clearly show that (except in the simulations of binding modes IV and VI) the mutation of Asp274 to alanine in the substrate-bound chain leads to an increase in the average distance and on the distance fluctuations of these residues relative to the non-mutated empty subunit.

Table 4: Distances (in Å) between the polar hydrogens in Tyr196 and Thr227 in wild-type and mutant coproporphyrinogen III oxidase.

Binding mode	Chain A		Chain B	
	Wild-type	mutant	Wild-type	mutant
I	6.56 ± 0.78	8.97 ± 1.21	5.42 ± 0.55	7.57 ± 1.10
II	5.50 ± 0.54	7.90 ± 0.64	5.40 ± 0.54	5.27 ± 0.46
III	5.52 ± 0.69	6.73 ± 1.11	6.16 ± 0.93	5.69 ± 0.72
IV	8.45 ± 0.99	7.98 ± 0.61	5.47 ± 0.33	5.50 ± 0.37
V	5.85 ± 0.75	7.73 ± 0.80	5.98 ± 0.63	5.43 ± 0.41
VI	7.70 ± 1.43	7.87 ± 0.36	6.08 ± 0.86	6.32 ± 1.23
VII	5.57 ± 0.46	7.02 ± 0.89	5.41 ± 0.54	5.56 ± 0.55

VIII	6.44 ± 0.62	9.03 ± 1.43	5.30 ± 0.46	6.48 ± 1.02
apoprotein	7.22 ± 0.59	n.d	5.69 ± 0.63	n.d

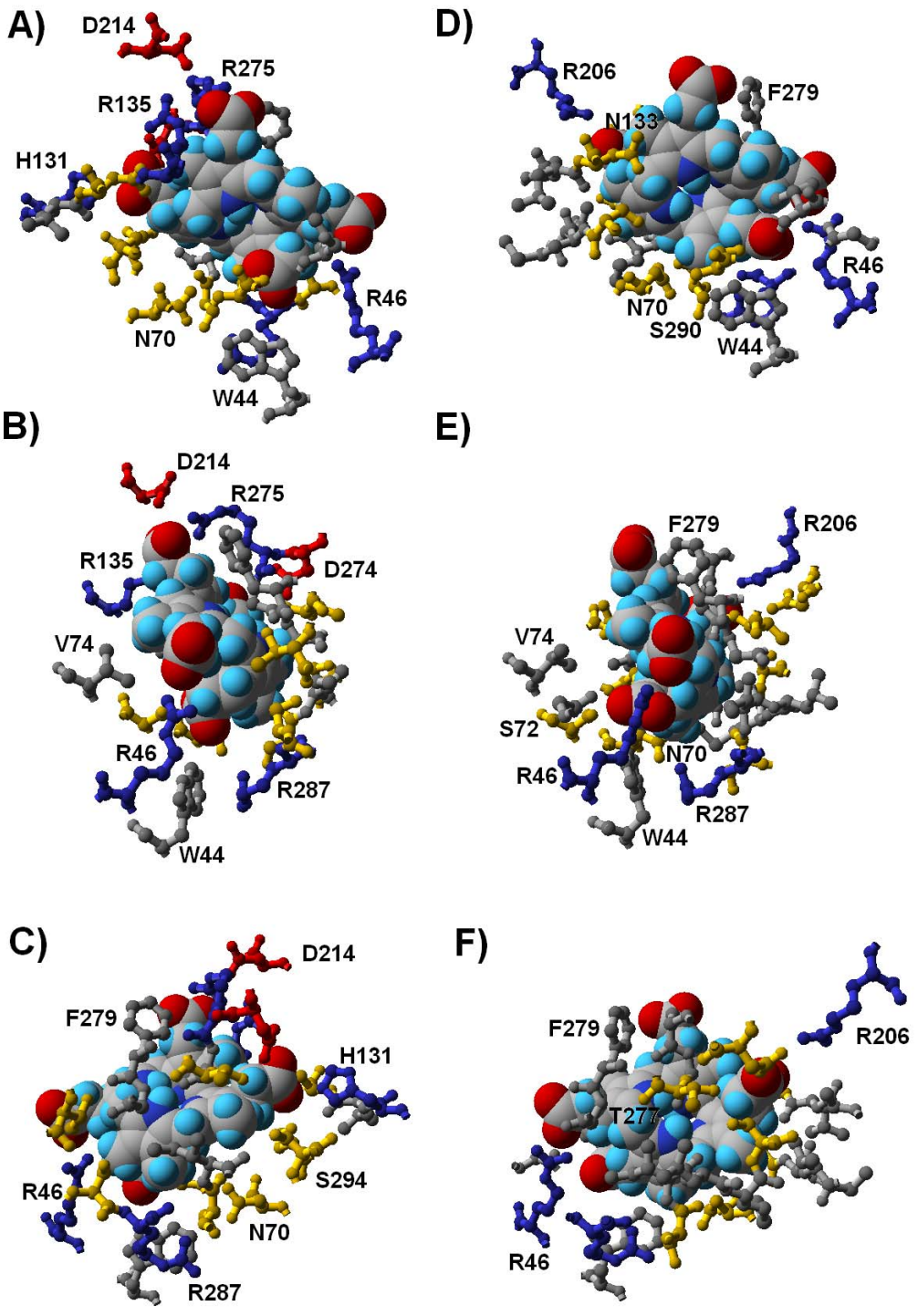


Figure 5: Detailed view of binding mode VIII. Panels A-C (wild-type-protein) are related to each other through sequential ≈90º rotations around the vertical axis. , as are panelsD-F (quadruple mutant)

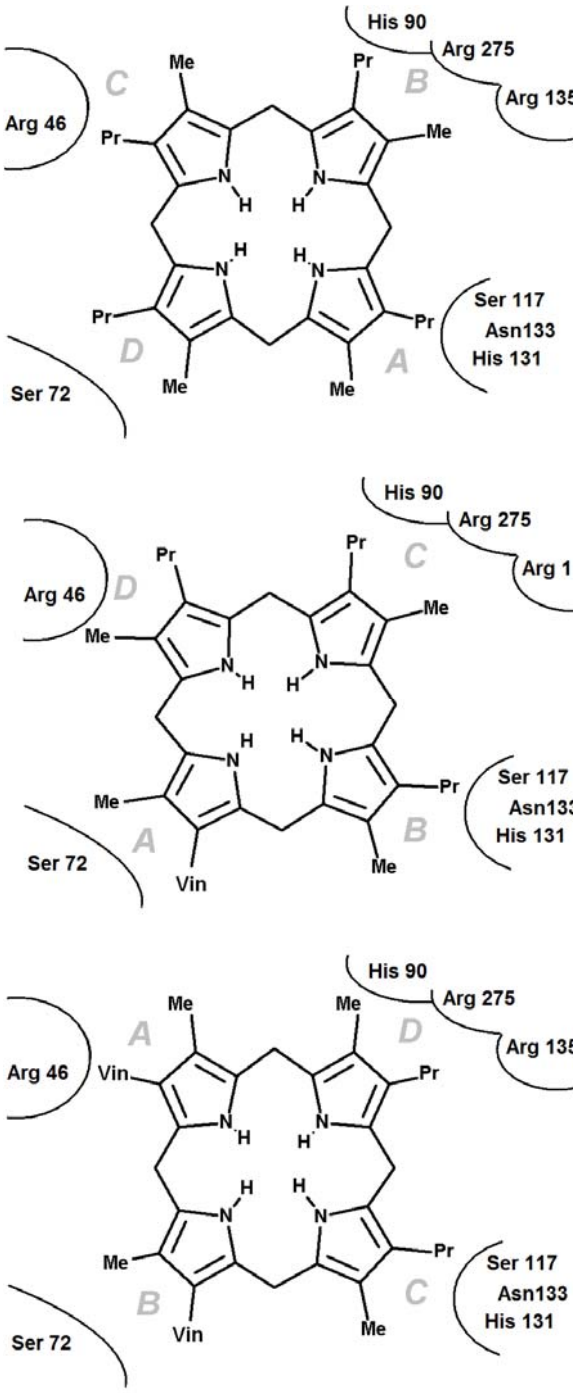


Figure 6: Schematic representation coproporphyrinogen III (top panel), harderoporphyrinogen III (center panel) and porotoporphyrinogen IX (bottom panel) to the enzyme active site

The binding pose obtained from VIII also explains why only rings A and B can be decarboxylated by the enzyme (Figure 6). It is known experimentally that decarboxylation first occurs in the active site region that holds ring A, which is shown by our simulations of binding mode VIII to be the region surrounding Asn133 and His 131. After the negative propionate in ring A is converted to a neutral vinyl group the favorable interaction with His131 disappears, and the binding affinity decreases. This partially decarboxylated intermediate (harderoporphyrinogen III) may leave the active site, and can only bind again in a new conformation (rotated ≈90º vs. the original one) with the vinyl group in ring A occupying the original binding site of the ring D methyl (which has enough room for a two-carbon substituent) (Figure 6, center panel). Ring C can easily occupy the binding site originally belonging to ring B, due to the identical distribution of substituents in these two rings, and ring D may occupy the original position on ring C (in spite of the opposite distribution of substituents) because this site is solvent-exposed and can therefore easily accommodate a propionate group in the place of the original methyl group. After the second decarboxylation, the binding affinity again decreases due to the loss of propionate-His131 interaction, and the product (protoporphyrinogen IX) leaves the active site. Protoporphyrinogen may not bind the enzyme again, as the only binding modes possible (Figure 6, bottom panel) are now less favorable since they depend on the existence of propionate groups on rings A and B (see description of binding modes IV and VI above). Furthermore, these binding modes also place the remaining propionate groups far from the detected active site (His131, Asn133 and Ser117), which renders catalysis ineffective.

Conclusions

In this study, the application of molecular dynamics simulations of enzyme in explicit solvent and periodic boundary conditions to the eight possible substrate orientations in coproporphyrinogen III oxidase active site cavity enabled the determination of the putative substrate binding mode and the extensive characterization of the enzyme active site, which was quite conclusive as far as the catalytic determinants of this particular reaction are concerned. In fact, the MD simulations described here enabled the determination of a substrate binding mode fully consistent with the known selectivity of the active site towards substituted tetrapyrroles, and explained the lack of activity of the H131A, R135A, D274A and R275A mutants and the reasons behind the non-occurrence of catalysis on the C and D rings of the tetrapyrrole. The finding of a strong hydrogen-bonding role for the invariant G276 residue in the successful binding mode offers a powerful argument against the reaction mechanisms that require the deprotonation of the reacting pyrrole ring, as this interaction is only relevant with protonated pyrrole nitrogens and the absence of cationic side chains close to the substrate pyrrole nitrogens prevents the protein from lowering the high pKa of pyrrole to values compatible with its deprotonation at physiological pH. Since the Arigoni mechanism (Figure 1A) is only feasible with a deprotonated pyrrole[18], these constraints leave the Lash mechanism (Figure 1B) with neutral substrate as the only mechanism compatible with the experimental and computational findings.

Supplementary Material Available: Coordinates of the enzyme-substrate complexes for wild-type and H131A/R135A/D274A/R275A quadruple mutants in the eight putative binding modes. Overview of the molecular dynamics simulations (RMSD, RMSF, distance between helixes H2 and H8, contacts between protein and substrate). Detailed views of binding modes I-VII in wild-type and quadruple-mutated coproporphyrinogen III oxidase.

V. References

1. Yoshinaga, T. & Sano, S. (1980) *J. Biol. Chem.* **255**, 4727–4731.

2. Medlock, A. E. & Dailey, H. A. (1996) *J. Biol. Chem.* **271**, 15765–15770.

3. Elder, G. H., Evans, J. O., Jackson, J. R. & Jackson, A. H. (1978*) Biochem. J.*, **169**, 215–223.

4. Yoshinaga, T. & Sano, S. (1980) *J. Biol. Chem.* **255**, 4722–4726.

5. Lash, T. D., Mani, U. N., Drinan, M. A., Zhen, C., Hall, T. & Jones, M. A. (1999) *J. Org. Chem.* **64**, 464–477.

6. Akhtar, M. (2003) in *The Porphyrin Handbook*, eds. Kadish, K. M., Smith, K. M. & Guilard, R. (Elsevier Science, New York), Vol. 12, pp. 69–86.

7. Cavaleiro, J.A.S.; Kenner, G.W & Smith, K.M. (1974) . *Chem. Soc. Perkin Trans. I*, **10**, 1188-1194

8. Sassa, S. & Kappas, A. (2000) *J. Intern. Med.* **247**, 169-178

9. Lee, D.-S., Flachsová, E., Bodnárová, M., Demeler, B., Martásek, P. & Raman, C. S. (2005) Proc. Natl. Acad. Sci. USA **102**, 14232–14237

10. Phillips, J. D., Whitby, F. G., Warby, C. A., Labbe, P., Yang, C., Pflugrath, J. W., Ferrara, J. D., Robinson, H., Kushner, J. P. & Hill, C. P. (2004) *J. Biol. Chem.* **279**, 38960–38968.

11. Gitter S.J., Cooper C.L., Friesen J.A., Lash, T.D., Jones M.A. (2007) *Med. Sci Monit.* **13**, BR1-10

12. Stephenson JR, Stacey JA, Morgenthaler JB, Friesen JA, Lash TD, Jones MA. (2007) *Protein Sci.* **16**, 401-410.

13. Elder, G.H.; Evans, J.O. (1978) *Biochem. J.*, **169**, 205-214

14. Porra, R. J. & Falk, J. E. (1964) *Biochem. J.,* **90**, 69-75

15. Al-Hazimi, H.M.G.; Jackson, A.H.; Knight, D.W.; Lash, T.D.(1987) *J. Chem. Soc. Perkin Trans. I*, 265-276

16. Lash, T.D.; Mani, U.N.; Keck, A.-S. I.M. & Jones M.A. (2010) J. Org. Chem., **75**, 3183-3192

17. Jones , M.A; He, J. & Lash T.D. (2002) J. Biochem., **131**, 201-205

18. Silva, P.J. & Ramos, M.J. (2008) Bioorg. Med. Chem. **16**, 2726-2733

19. Lash , T.D. (2005) *Bioorg. Med. Chem. Lett.* **15**, 4506-4509

20. Krieger E, Darden T, Nabuurs S, Finkelstein A, Vriend G (2004) *Proteins* **57**, 678-683

21. Wang J, Cieplak P, Kollman PA (2000) *J. Comput. Chem.* **21**, 1049-1074

22. Essman U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) *J. Chem. Phys.* **103**, 8577-8593.

23. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. (1984) *J. Chem. Phys.* **81**, 3684-3690.

24. Jakalian A, Jack DB and Bayly CI (2002) *J. Comput. Chem.* **23**,1623-1641

25. Krieger E, Nielsen JE, Spronk CA, Vriend G (2006) *J. Mol. Graph. Model.* **25**,481-486

26. Shoolingin-Jordan P.M. (2003) The Biosynthesis of Coproporphyrinogen III. In Kadish, K.M., Smith, K.M. and guilard, M. (eds.) *The Porphyrin handbook. Volume 12: The iron and Cobalt Pigments: Biosynthesis, Structure and Degradation.* Elsevier Science

27. Silva, P.J. & Ramos, M.J. (2008*) J. Phys. Chem. B*, **112**, 3144-3148