Identification of a New Site of Metabolism for Phenprocoumon by Modeling it's CYP2C9 Hydroxylation Pattern

Quiroga I[1], Melendez FJ[1], Salvador-Atonal K[1], Kammerer B[2] and Scior T[1*]

[1]Faculty of Chemical Sciences, Benemerita Autonomous University of Puebla, C.P. 72570 Puebla, Pue., Mexico, State Puebla, Mexico
[2]Center for Biosystem Analysis ZBSA, University of Freiburg, Freiburg, Germany

*Corresponding author: Scior T, Laboratory of Computational Molecular Simulations, Building 118/FCQ3, University City, Benemerita Autonomous University of Puebla, C.P. 72570 Colonia San Manuel, City of Puebla, Mexico.Tel: 52.222.2295500 Ext. 7529, E-mail: tscior@gmail.com; thomas.scior@correo.buap.mx

Abstract

Molecular modeling was conducted to study the observed sites of metabolism for cytochrome P450 enzyme CYP2C9 and find a hitherto unknown hydroxylation site for phenprocoumon. Four hydroxylation sites of phenprocoumon metabolism by CYP2C9 were taken from our previous site of metabolism (SoM) identification by mass-spectrometric and selected-reaction-monitoring chromatographic analyses. Phenprocoumon (PPC) and warfarin (WFN) were successfully docked into CYP2C9 and CYP3A4. Docking results for the four monohydroxylation metabolites (6-OH-PPC, 7-OH-PPC, C2'-OH-PPC and C4'-OH-PPC) showed final docking poses in good keeping with the experimental data. The catalytic cycle was revised and ternary coordination complexes between heme B, reactive oxygen and substrate with a tetrahedral geometry (atoms Fe-O-CH) were identified under the molecular mechanics force field (Yeti). Additional quantum mechanics calculations (Gaussian09) also focused on the ternary coordination complex (iron IV - activated oxygen - sp3 hybrid carbon atom). Transition state and equilibrium structures for OH-Ph-S complexes of iron porphyrin moiety were computed with density functional theory (DFT) at B3LYP level of theory. To establish the most preferred mode of interaction between the complexes and the protein structure, the DFT calculations have been used to accomplish the molecular docking studies. As a direct result the hitherto unknown structure of a PPC metabolite could be identified as the 10-hydroxy-metabolite of PPC. To simulate the hydroxylation of the ortho position (2') at the phenyl ring, however, a customary model (rotamer library) was generated. After rotation of three binding-relevant side chains (Val113, Leu365 and Thr301) the sterically congested 2' position became accessible to enzymatic hydroxylation.

Keywords: 4-hydroxycoumarin; Vitamin K; Warfarin; Docking pitfalls: *Ab inito*

List of abbreviations: PPC: Phenprocoumon; SRM: selected-reaction-monitoring (chromatogram); TS: Transition state; MS: Minimum state; WFN: Warfarin

Introduction

Acenocoumarol (ACE), phenprocoumon (PPC) and warfarin (WFN) are commonly used for the causal or prophylactic oral treatment of venous and arterial thrombosis (vitamin K–dependent blood clotting). Receivers of oral anticoagulants show a large variability in bleeding risks. Bleeding is a side effect of hospitalized patients on anticoagulation therapy which occurs with a frequency between 15 to 20 percent [1]. Impaired biotransformation and elimination of frail patients may contribute to the variable bleeding risk known as hemorrhagic risk by coumarin anticoagulants. The molecular target of these anticoagulants is the vitamin K epoxide reductase complex [2]. These drugs undergo vast biotransformation in the liver by members of the microsomal metalloenzyme family known as cytochrome P450 oxygenases (CYP for short). Our target enzyme CYP2C9 as well as other members participate in phase 1 biotransformation of body substances or xenobiotics (drugs) by demethylation, oxidation, or hydroxylation reactions [3]. Prior to our computational molecular simulation study, the relative human liver microsome contents of CYP2C9 were assessed against other CYP liver enzymes (CYP1A2, 2A6, 2C19, 3A4) by means of spectrophotometric measurements by a group member [4-8] (Figure 1). In particular five monohydroxylated phenprocoumon metabolites were identified (positions 6, 7, 8 2' and 4') in human plasma by liquid chromatography mass-spectrometric coupling [6]. Our present *insilico* study aims at modeling those five

experimentally identified hydroxylation sites of phenprocoumon (figure 2 in [6]). In view of the enzymatic mechanism based on the heme group at the active site of all CYP enzymes and the fact that all monohydroxylated metabolites – except for one – were analytically identified, we also aim at delivering the long-awaited answer to the unsettled question whether the hitherto unidentified peak in the chromatogram could be ascribed to one of the remaining three structural positions at either C9, C10 or C11 (figure 2 in [6]). Thus, the additional goal was to computationally identify the missing hydroxylated PPC metabolite (C9 or C10 or C11-OH-PPC) [4-8].
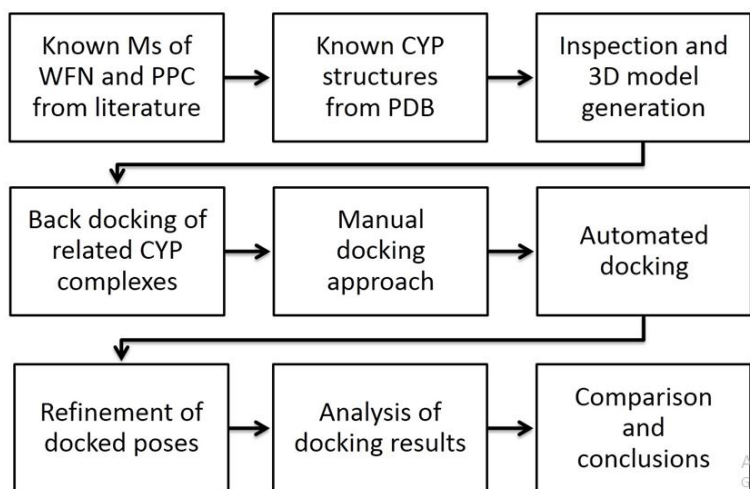

Figure 1: Workflow of the docking study with the 9 successive stages (cf. 9 boxes). Swiss PDB viewer, VEGA ZZ and FlexX under Sybyl were used at stages 3, 4 and 5. At stages 6 on 7 a combination of Autodock 4 and docking program Yeti 8 with Biograf R were used in a more systematic way. Abbr. "Ms" for metabolite(s)
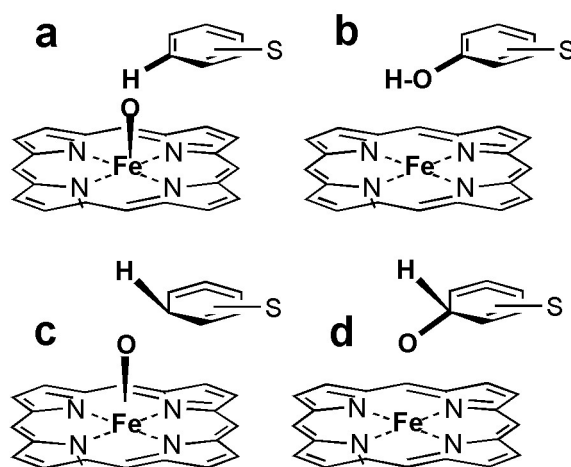

Figure 2: The four options (a through d) of possible molecular geometries and the atom types required to formulate the input for docking. While options a and b have an intact aromatic ring with sp2 hybridization and aromatic bond types, options c and d operate on a carbon atom with sp3 hybridization in tetrahedral bond geometry. Option d circumvents the necessity of replacing a (undefined) coordination with (parameterized) covalent bonding, and proofed superior results during docking

To our best knowledge substrate coordination to a multi-step catalytic cycle is a nontrivial endeavor. This way, we had also to unveil which catalysis step is amenable to a molecular mechanics formalism (docking algorithms), and whether standard docking tools encompass the atom types of the catalysis configuration (intermediates, transition states) and therein be applicable to substrate - enzyme docking (coordination complex). Another not less worrying challenge was to face the renowned induced fit mechanism since standard docking means coupling a flexible ligand to a rigid receptor. In particular, this induced fit phenomenon occurs upon ligand - receptor complex formation and consists of conformational changes of one or both molecular partners. They can comprise smaller subtle changes in torsional angles at side chains, or even main chain movements on the proteins, and entire domain rotations at a large scale of multi-subunit, multimeric complexes. To further the reader's interest, three reviews lend detailed insight into this topic of protein flexibilities [9-13].

Molecular docking is the mainstay computational technique to describe the binding between a ligand to its target receptor protein. Each ligand pose is evaluated by a score to reflect its binding affinity [14]. It is, however, not straightforward to use any conventional

docking tool to compute the phenprocoumon (PPC) binding to the heme group at the catalytic site which forms a monodentate coordination between the metal center and the substrate intermediate during the oxidative biotransformation step in the liver cells [15].

To this end we combined molecular mechanics (ligand docking to receptor) and quantum mechanics (*ab initio* transition state calculations of the coordination complex) to generate proper molecular models of the sites of metabolism (SoM) upon PPC binding to the active site of CYP2C9.

Almost all drugs are noncovalently bound to their receptors. But in rare cases a reversible or irreversible covalent bond is formed between them (e.g. inhibitors like carbapeneme of penicillinase or parathion of cholinesterase). Hence, new docking techniques have been evolved for covalent docking (e.g. Covdock, Covglide). In a seminal work by A. Bender and coworkers, the software which can predict CYP activities was reviewed [16]. Moreover, the more general concept of ligand - receptor docking has already been applied to the particular field of substrate - CYP enzyme metabolism (Table 1). Thereby, the most frequently used programs include Autodock 4, AutodockVina, GOLD and FlexX [17-20].

Year	Description and key words	References
2006	Overcoming the receptor's rigidity of its backbone by assessing its structural flexibility during **molecular dynamics** (MD) studies with explicit water. But MD with explicit solvent simulation has two mayor flaws: (i) the limited computational resources and (ii) the addition of even tiniest calculation errors (oversimplifications by force fields, parameters and electronic and entropic effects) into relevant model artifacts during large time periods of simulations during tens of nanoseconds.	[21]
2007	Combining softdocking from receptor ensembles and MD creates a new **approach called Fleksy.**	[22]
2007	To circumvent the missing conformational flexibility during rigid body docking the so-called **soft-docking** approach can be applied. It reduces the impact of the repulsion term in the force field equations between ligand and binding site.	[22]
2008	The application of grid-based docking approaches was reviewed in a seminal work.	[23]
2009	Defining a few binding-relevent amino acids at the binding site to simulate conformational changes of their side chains. For example Autodock 4.2.	[17]
2010	Use of Linear Intereaction Energie (LIE) with Merck Molecular Force Field and ligand-based virtual screening fot identification of CYP1A2 ligands by structure-based	[24]
2011	The proposition of an *in silico* method that predict the SoMs as "metabolophoros"	[15]
2013	Docking with GOLD was used to describe the impact of CYP1 inhibitors on regioselectivity for hydroxylation reactions.	[25]
2015	Grid docking by Glide software was combined with 3D QSAR techniques to elucidate the CYP2C19 inhibition activity of certain alkaloids.	[26]
2015	A special case is the simulation of covalently bound ligands to form receptor complexes.	[27]
2017	Using many different crystal structures of the same protein (related PDB entries) for docking to describe protein flexibility. This is known as docking by **multiple receptor structures**, i.e. a receptor ensemble is used as target to reflect the protein flexibility, e.g. Glide ensemble docking.	[28]

Table 1: Listing of published concepts for ligand - receptor docking and examples applied to CYP biotransformation during the last decade

Modeling of the hydroxylation pattern

Computing the site of hydroxylation for the structurally undefined OH-PPC metabolite seems a reasonable endeavor since molecular modeling techniques have already been applied to the field of CYP metabolism (figure 2 in [6]).

The complex affinities were assessed by ligand - receptor docking simulations (Sybyl FlexX, Autodock 4 and Yeti 8) [17,20,29]. This process engages the sixth coordination bond, the electron movements of which cannot be directly computed using any molecular mechanics (docking algorithms). Hence, it was treated separately by an *ab initio* approach (Gaussian software package) [17,30,31].

The crystal structures were downloaded from PDB, their superposition conducted under Swiss PDB Viewer, and manual docking was carried out under VEGA ZZ [32-34]. Automated docking was performed by FlexX (Sybyl X) and Autodock 4 (MGL tools) [17,18,20,35]. The final poses were treated under Yeti 8 and Biograf R [36,29]. We point out, the rationale behind the decision to refine local minima of the final geometries lies in Yeti's implementation of specific parameters for iron and heme groups. Precisely, it constitutes an extension to the AMBER force field [36]. Related crystal structures of ligand-CYP enzyme complexes (coumarin ligand with PDB code: 1Z10 and flurbiprofen with PDB code: 1R9O) were inspected and back docking of the extracted ligand evaluated to validate the blind docking of PPC into human CYP2A6 or CYP2C9 [37-38].

Methods for computational docking simulations

For decades docking tools have focused on the noncovalent interactions, but we cannot neglect the pivotal role of heme at the active site. Between enzyme and its substrate there exist potential energy contributions from - not only noncovalent interaction – but also complex coordination. While the former are built-in components of parameterization sets of conventional docking algorithms, the latter are nonstandard features [17,29]. A schematic workflow of nine steps was devised to combine manually docked poses as start positions for automated docking by Autodock and refinement of final poses by Yeti 8 (Figure 1) [17,29].

Results

Mass spectroscopic evidence provides a strong indication that the phenyl ring of the substrate is hydroxylated at the 2' (two prime) position. The final docking poses showed a ligand orientation similar to the crystal structure of flurbiprofen with significant resemblance to structural features of PPC (flurbiprofen, PDB code: 1R9O) [6,8,38].

After comparing their structures we carried out selfdocking to assess which step of the catalytic cycle is suited as the final pose (docking endpoint) and how to coin this substrate - enzyme contact into the molecular mechanics formalism for Autodock 4 and Yeti 8 (Figure 2).

In this work, a theoretical study of the mechanism of reaction between the heme group and the aromatic ring is presented, using DFT-B3LYP method. A model of the general mechanism is proposed for this reaction in gas phase (Figure 3).
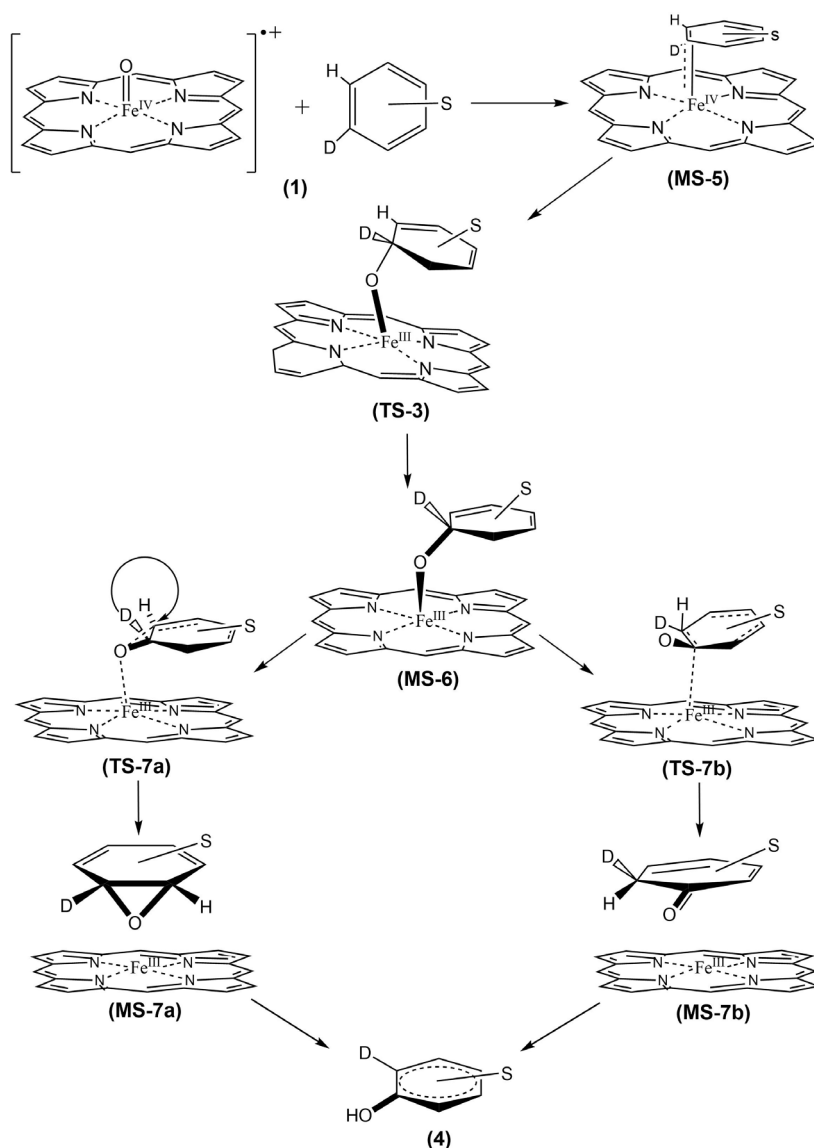

Figure 3: The substrate (S) with its aromatic ring approaches the catalytic site with the heme group to form a tetrahedral transition state (2). A ternary coordination complex is formed and embraces the heme iron IV and the activated oxygen moiety which in turn is bound to an sp3 hybrid carbon atom (**TS-3**). Two reaction pathways are possible (**TS-7a** and **TS-7b**) the hydroxylated metabolite [39]. The mechanism is known as "NIH shift", i.e. a hydrogen atom is replaced by deuterium (D) to reflect the atom shift. The docked ligand poses adopted a sp3 hybrid geometry of the attacked carbon atom in order to simulate the reaction transition state into the product (**4**)

In the first step, from (MS-5) structure, it is proposed the formation of a tetrahedral transition state (TS-3) for obtaining the intermediary (MS-6). This intermediary has the structure like a ternary radical cationic complex formed by a central carbon atom with temporary sp³ hybridization: a type of reactive oxygen and the cation of iron as a catalyst with charge +III (Fe^{III}) on the side of heme group. The second step of the reaction is the posible formation of two species (MS-7a and MS-7b): one corresponds to the formation of an epoxide group and another of a keto group, both from their respective transition state structures (TS-7a and TS-7b). These last species are completely dominated by electronic influence in order to stabilize the corresponding intermediaries (MS-7a and MS-7b). Finally, in the case where the alcohol is only involved as final product (4), an hydrogen atom was replaced by deuterium (D) to analyze the mechanism known as "NIH-shift", whose final product is of interest in this study for molecular docking simulation [39].

Regardless of the subtitle (quantum) differences between both reaction coordinates, the aforementioned tetrahedral ternary complex constitutes the point of departure in common for both. Hence, that configuration is chosen as a static end point of docking and as a predefined transition state (TS) for *ab initio* calculations.

Best ranked docking solutions were extracted from the 2.5 Å RMSD clusters (stage 6) for binding energy refinement (stage 7). At stage 7 (Figure 1) the docked ligand complexes were refined under the Yeti force field. It is an AMBER extension and includes parameters for ion and heme [29]. Thereby, all final poses obtained a natural octahedral configuration and were ranked under Yeti 8 [29,40]. At stage 8, inspection and evaluation of the final poses took place. The accepted docking solutions were compared to the binding geometries observed in the crystal structures of hydroxylated camphor CYP P450 complex [41,42]. The x-ray structure of CYP2A6 with co-crystallized coumarin was used as another 3D-template because it revealed the geometries between substrate and the heme - oxygen system (PDB codes: 1Z10, 2CPP and 3CPP) [41,42,37]. Coordination geometries were measured: Fe-O-C atom distances, Fe-O-C angle, and dihedral angles between the heme group and docked ligands (Figure 4) [43].
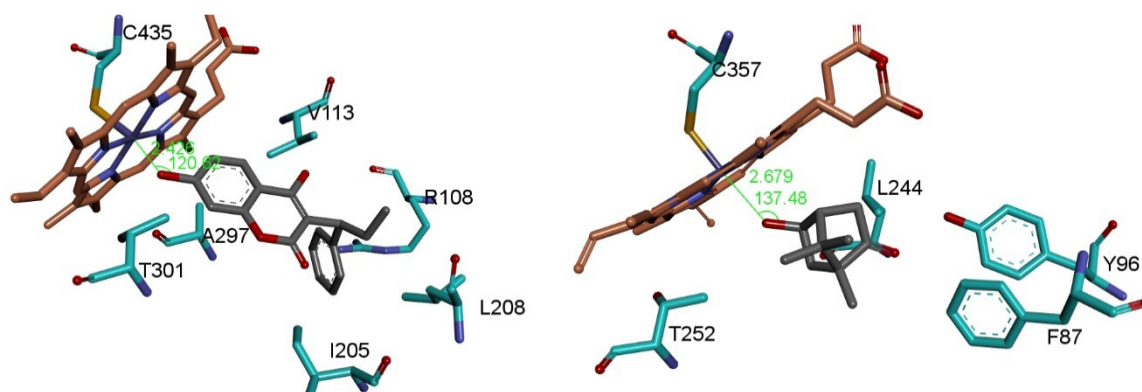

Figure 4: Geometrical comparison between final poses of 7-hydroxylated PPC (left) and the observed case of hydroxylated camphor complexes (PDB codes: 3CPP, 1NOO) [59,42]. Color code: O red, N blue, C grey, Fe orange, S yellow, H white [43]. The views show each ligand in the center and the heme group leftmost on top of the ligands

No crystal complex with PPC exists, but with structurally related WFN which was co-crystallized with a mutant type CYP2C9 (PDB code: 1OG5) [44,45]. So we decided to study WFN, too. With the catalytic site nearby the mutated side chains, steric hindrance caused the substrate displacement further away from the heme group. Finally, the refined poses from Yeti were found plausible upon comparison with related crystal complexes (Figure 5) [38,41,42,44-47].

In the case of PPC C2' we found indications for induced fit with docked solutions reflecting conformations other than the available crystal structures (Figure 5). At stage 9, we compared the observed chromatographic data to the calculated affinities for correlation between the amount (AUC) and the potential energies of docked poses (Table 2) (figure 2 in [6]). Upon analyzing the binding energies of the final poses it can be seen to illustrate one instance that C7 is energetically preferred over C6 hydroxylation. This is in good keeping with the SRM chromatogram (figure 2 in [6]).

The hydroxylation pattern of the main metabolites (see circle 1 = 4'-OH-PPC; circle 2 = 2'-OH PPC; circle 4 = 6-OH-PPC; circle 5 = 7-OH-PPC) were successfully reproduced except for circle 2 due to steric hindrance. This is a hint for shifting side chains or backbone (induced fit) which cannot be handled adequately by rigid protein docking procedures (Figure 2) (Table 2) [6]. As source of the steric clashes we identified Val113, Leu365 and Thr301 on CYP2C9 (PDB code: 1R9O, resolution 2 Å, R value 0.2) [38]. To avoid bad substrate - enzyme contacts (clashes) new conformations for their side chains were selected from a rotamer library to allow the 2' hydroxylation [33,49].

The spectral data could also be reproduced in a semi-quantitative way: the (S)-enantiomer always fitted better into the catalytic site which explains why the (S)-metabolites always appears under a smaller area under curve (AUC) than the (R)-metabolites of PPC. The biotransformation rate of the (S)-forms should then be higher to explain why the blood (plasma) concentration was smaller. This parallels the finding that CYP2C9 is the main metabolizer of S-WFN but only marginally of R-WFN substrate [6,50].
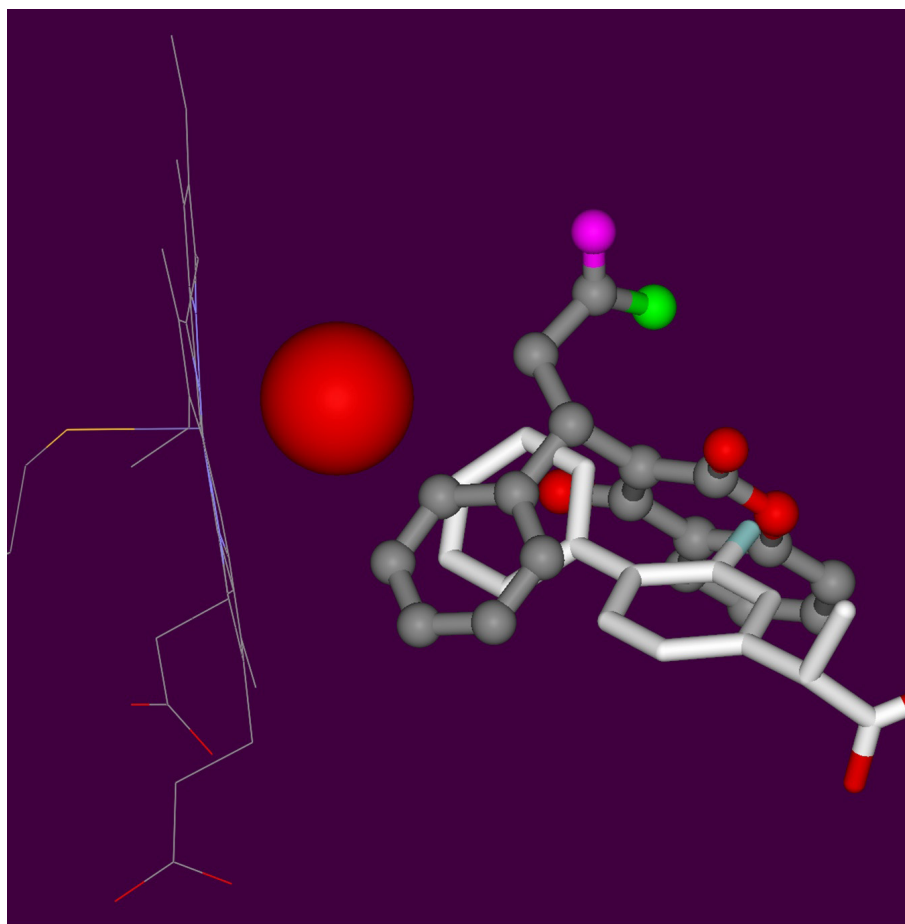
Figure 5: Geometrical comparison between final poses of 2'-hydroxylated PPC and the observed case of hydroxylated flurbiprofen (white carbon atoms) complexes (PDB codes: 1R9O) [38]. Color code: O red, N blue, C grey, F light blue, Fe orange, S yellow, H omitted for better viewing. The two heavy atoms which make the distinction between warfarin and phenprocoumon are colored in magenta (O=) and green (-CH3). The oxygen species attached to heme iron is presented as a sphere with full atom radius. The proximal cysteine (Cys435) is displayed, too (leftmost yellow stick). The observed distances: Fe - hydroxyl O (2.4 Å), Iron Fe-C atom of SoM (4.9 Å), hydroxyl O–C atom of SoM (3.0 Å) [101]. The required coordination distances had the same lengths as found in the crystal structures: Fe-O between 1.6 and 2,4 Å or O—C2' between 3.0 and 3.1 Å, respectively [38,42]

PPC2′	Total E [Kcal / mol]	Electro-static IA	VdW IA	H-bonding	Metal IA	Total E minus Mtl IA
a	-9	-1	-8	-0.1	0	-9
c	-9.5	-1.5	-8	-0.1	0	-9.5
d	-20	-0.01	-7	0	-13	-7
b	-19	-4	-1	-0.1	-13	-6
PPC4'						
a	-7	0.2	-7	0	0	-7
c	-7	-0.1	-7	0	0	-7
d	-30	-7	-6	0	-16	-14
b	-29	-8	-6	-0.3	-15	-14
PPC6						
a	-3	1	-3	-0.002	0	-3
c	-3	-0.1	-3	-0.002	0	-3
d	-24	-5	-4	-0.004	-16	-8
b	-41	-18	-6	-2	-27	-14
PPC7						
a	-0.7	1	-1	-0.003	0	-0.7
c	-7	-0.002	-7	-0.001	0	-7
d	-37	-10	-2	0	-26	-11
b	-34	-14	6	-0.01	-26	-8

PPC2′	Total E [Kcal / mol]	Electro-static IA	VdW IA	H-bonding	Metal IA	Total E minus Mtl IA
PPC10						
a and c	-21	0	-21	0	0	-21
d	-24	0	-24	-0.005	0	-24
b	-24	0	-24	-0.004	0	-24
FLP						
a	-31	-18	-9	-5	0	-31
c	-31	-18	-9	-5	0	-31
d	-48	-25	-7	-3	-14	-34
b	-41	-24	-3	-2	-12	-29

Table 2: Listing of docking solutions at the catalytic site. 1st column refers to the four options (a through d) of possible molecular geometries (Figure 2). The hydroxylation site was accessed through the oxygen-iron (III)-heme-Cys435 octaeder. In the case of circle 3 as the missing structural datum, all 4 remaining possibilities (not hydroxylated sites) were evaluated (9R10S, 9R10S, 9S10R, 9S10S). "No" implies necessary spatial changes which reflects induced fit, e.g. side chain rotations on Val113, Leu365 and Thr301 made the 2' position accessible to hydroxylation [49]

Yet these findings leave one question still to be answered concerning the peak labeled as circle 3 in the spectral analysis (figure 2 in [6]). To complicate matter, early attempts in docking PPC for the ortho-position (2') on its phenyl ring fell short of expectation. Probing with WFN it was found that the ortho-position (2') on the phenyl ring actually was accessible to oxygenation (Table 2). The apparent contradiction was resolved acknowledging that PPC was docked into the empty active site of the crystal structure liganded to flurbiprofen [38]. Apparently the ligands invoke induced fit phenomena leading to structural rearrangements between PPC substrate and enzyme (spatial requirements). Since no crystal structure of PPC - CYP2C9 has been elucidated, we simulated the missing induced fit by rotamer changes (Val113, Leu365 and Thr301). This way C2' hydroxylation of PPC was successful. This finding extends the known pattern of metabolism which we summarized in the introduction chapter. Playing back the final poses for PPC and WFN, position C9 on both were clearly an unlikely place for hydroxylation because the C9 ramification lead to spatial congestion shielding against catalytic attack (Table 2).

The understanding of steric requirements as the essential driving forces for regioselectivities was furthered by two facts from literature: (i) C9 is not metabolized on WFN and (ii) the more accessible position C10 of WFN is hydroxylated instead. In good keeping the C10 hydroxylation of PPC was the most populated solution cluster after scoring the docked poses (Table 2). On theoretical grounds it was concluded without ambiguity that the peak without structural information labeled as circle 3 be ascribed to C10-OH-PPC. C10-hydroxy-WFN was found to be the second most abundant plasma metabolite (Figure 2) [51].

Results for the computational transition state simulations

Our *ab initio* study was carried out to generate possible transition states by quantum mechanics.

Albeit, it does not lie within the scope of this study to evaluate the extant literature on heme enzyme docking it is save to utter that this study explored all four options to formulate the docking start geometries to reflect the catalytic circle on both, a molecular mechanics level defining adecuate atom types and on quantum mechanics level, too (Figure 2).

In order to analyze the electronic and geometrical propensities used as starting geometries of the molecules for docking studies, the reaction mechanics was studied by a quantum mechanics approach. The structures of reactants, transition states and products were optimized. To this end only functional parts of the larger molecular entities were used, so that the fragments were amenable to "ab initio" treatments (Figure 6 and 7).

A hybrid-GGA functional, named B3LYP, was used for carrying out the calculations. Hybrid functionals mix exchange energies calculated in an exact (Hartree-Fock-like) approach with those obtained from DFT methods in order to improve the task performance. Several analyses on the assessment of the B3LYP performance have been made for interactions in bio-molecules and clusters which confirm the accuracy of this method [52-54].

The optimization for the present quantum mechanical study was performed at the B3LYP level using the 6-31G basis set applied to C, H, O and N atoms [55]. A relativistic effective core potential (ECP) on Fe atom was used to replace the inner core electrons, keeping the outer core electrons. In addition, the harmonic frequency calculations and zero-point vibrational energies (ZPE) were calculated at 298.15 K without considering any scale factor using the same theory level [56]. The free energies (G) of the reactants, intermediates, transition states, and products were obtained using the equation $G = E_{total} + ZPE - TS$, where E_{total} is the total energy of the species, T is the temperature in Kelvin and S is the entropy. All calculations were carried out using the Gaussian09 program [30]. The optimized geometries structures were visualized using GaussView 5.0.8 graphical application [57].
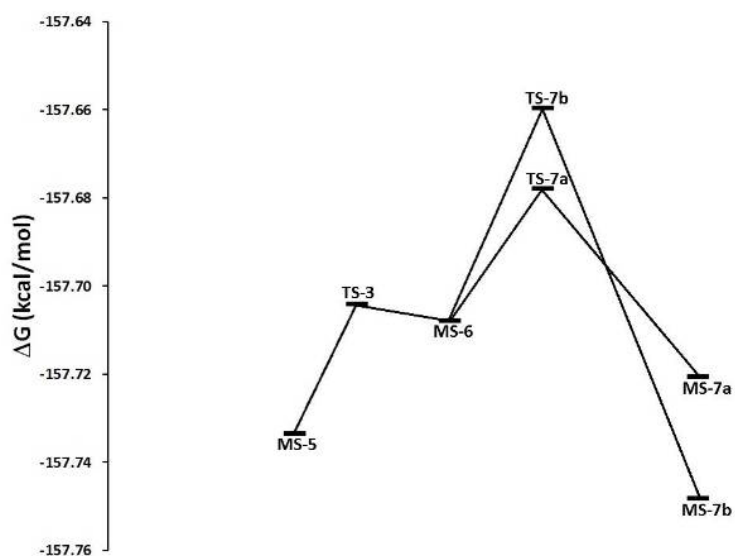
Figure 6: Minima (MS-n) and transition state structures (TS-n) located on the potential surface for the catalytic reaction (Figure 7)
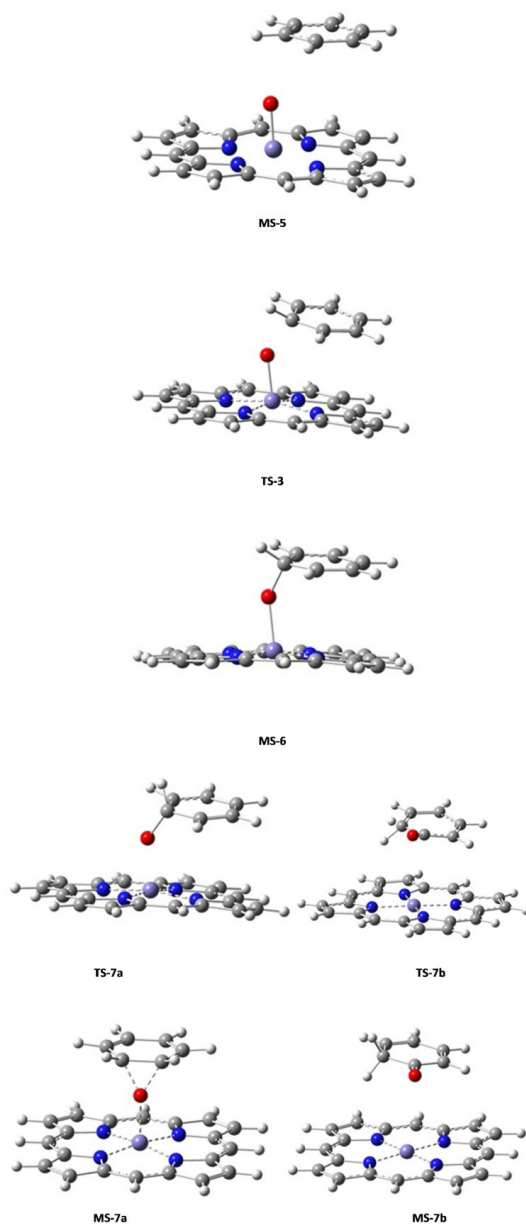

Figure 7: Potential energy surface profile for the reaction. Structures of minima and transition states are indicated in accordance with Figure 6

The structure and electronic energy of iron-porphyrin (FeP) system were calculated in gas phase. Optimized geometries for three oxidation states ($Fe^{(II)}$, $Fe^{(III)}$, and $Fe^{(IV)}$) and different multiplicities for iron were evaluated (Table 3). According to the obtained results, the most stable specie corresponds to $^3Fe^{(II)}$. The following trend in relative energies, in kcal mol^{-1}, is observed from this structure: $^1Fe^{(II)} < {}^5Fe^{(II)} < {}^4Fe^{(III)} < {}^2Fe^{(III)} < {}^6Fe^{(III)} < {}^5Fe^{(IV)} < {}^1Fe^{(IV)} < {}^3Fe^{(IV)}$. A value of 12 kcal mol^{-1} between the most stable specie and $^1Fe^{(II)}$ is found, while concerning $^3Fe^{(IV)}$ the relative energy is of 737.75 kcal mol^{-1}.

	$Fe^{(II)}P$		
	Triplet	**Singlet**	**Quintent**
E_{relat}	0.00	12.88	17.14
F_e–N_p	2.01	1.99	1.97
$Fe^{(II)}$	0.738	0.737	0.971
N_p	-0.655	-0.621	-0.710
	$Fe^{(III)}P$		
	Quartet	**Doblet**	**Sextet**
E_{relat}	321.46	326.63	333.62
F_e–N_p	1.98	1.97	1.98
$Fe^{(III)}$	1.028	1.026	1.013
N_p	-0.703	-0.706	-0.701
	$Fe^{(IV)}P$		
	Quintent	**Singlet**	**Triplet**
E_{relat}	721.87	737.29	737.75
F_e–N_p	1.96	1.99	1.99
$Fe^{(IV)}$	1.065	0.857	1.064
N^p	-0.696	-0.615	-0.695

Table 3: Relative energies regarding the ground state (E_{relat}) in kcal/mol optimized geometrical parameters (Fe–Np) in Å and Mulliken atomic charges of Fe and NP atoms corresponding to the lowest-lying spin state of $Fe^{(II)}P$, $Fe^{(III)}P$ and $Fe^{(IV)}P$ at B3LYP/6-31G level plus the Land2DZ basis set were used for the Fe atom (only 10 inner core electrons were described with Land2DZ pseudopotential)

An important role in the applications of quantum chemistry calculations for molecular systems is the calculation of Mulliken charges, because the charge on atoms affect to properties as electronic structure, molecular polarizability and dipole moment, among others. For example, the distributions of charge on atoms forming molecules suggest the formation of acceptor or donor pairs involving charge transfer. In addition, the atomic charge analysis can be used for describing electronegativity and it is also useful for charge transfer processes in chemical reactions. The results of Mulliken charge densities were also calculated using the B3LYP method.

An interesting correlation between the charge and the interatomic bond Fe–Np was found, being Np the nitrogens of the porphyrin ring. When a major charge exists in the system, a decreasing interatomic bond Fe–Np is obtained. The Mulliken atomic charges show, in all the cases, that the iron atom has a positive charge, while the charge on nitrogen atom is negative. These results demonstrate that when the iron atom changes of Fe^{II} to Fe^{IV} (major charge in the system) the charge is major on iron atom. So, the nitrogen atom has a value of maximum negative charge close to -0.710 in the quintuplet state of $Fe^{II}P$. The maximum positive charge is obtained for the Fe in the quintuplet state of $Fe^{IV}P$, whose distance Fe–Np corresponds to the small value found, showing the charge-distance relationship for the different systems studied in this work.

As a most valuable asset, our theoretical results obtained here are in good agreement with experimental solid state results and previous theoretical calculations obtained by BLYP/DPN level of theory [57,58].

Concerning the reaction mechanism the potential energy surface (PES) profiles assessed the energy minima, intermediary and transition state geometries. Three main transition states (TS-3, TS-7a and TS-7b) were found for the second stage of the hydroxylation mechanism (Figure 3). At that stage, the benzene moiety started its approach and the iron complex formed an intermediate (Figure 7). The transition state called TS-3 then showed an activation free energy of -157.7042 kcal mol^{-1} which fairly lay above the value of intermediate MS-5. Then the complex iron-porphyrin-O-phenyl (MS-6) was formed with a free energy of -157.7080 kcal mol-1 (Figures 3 and 6). In this step, about -0.0038 kcal mol^{-1} was delivered upon complex formation, all of which reflected the feasibility to form this complex as an intermediate (Figure 7). Prior to our work, two ways toward the production of the final compound had been experimentally observed (Figure 7) [39]. Both were amenable to model simulation during the third step: the migration of the Deuterium atom (atom D replacing atom H to observe the NIH shift) between two carbon atoms in both cases allows the complex formations MS-7a as well as MS-7b. The migration occurs through the TS-7b formation, with an activation free energy of -157.6596 kcal mol^{-1} at the expense of -0.0484 kcal mol^{-1}. Once migration has been achieved, the

final alcohol product can be obtained from TS-7a, with an activation free energy of -0.019 kcal mol^{-1} what fairly lies below the energy level of TS-7b with an energy consumption of -0.0299 kcal mol^{-1}. For this last step of product formation, a mechanistically plausible intermediate (MS-7a) was calculated and its formation energy below the intermediate (MS-5) was assessed: -157.7206 kcal mol^{-1} which reflects the energy release of 0.0887 kcal mol^{-1}. All values of formation and activation free energies were calculated and compared relative to the ground state (Table 3). The respective geometries are taken into account to formulate the discrete geometries (Figure 7). The Figure 7 shows the minima and transition state structures. Three main transition states (TS-3, TS-7a and TS-7b) were found in the second step of the mechanism. In this step, the biomolecular process begins with the approach of the benzene molecule and the complex FePO, forming the intermediate (MS-5). The transition state TS-3 (Figure 6) showed an activation free energy of -157.7042 kcal mol^{-1}, while the complex iron-porphirin-O-phenyl (MS-6 in Figure 6) was modeled to assess the free energy of -157 kcal·mol^{-1}. The formation of the complex MS-7b is obtained from the transition state TS-7b, with an activation free energy of -157.6596 kcal·mol^{-1}. On the other hand, the alcohol formation can be obtained from the TS-7a, with an activation free energy of -0.019 kcal·mol^{-1} below TS-7b (Figure 3). In this last process, a possible intermediate (MS-7a) is calculated, with a formation energy of -157.7206 kcal mol^{-1}. The values of formation and activation free energies are shown in Table 4. The respective geometries are depicted in (Figure 6 and 7). The discrete geometries are displayed in Figure 7.

Structures	ΔG	ΔG#	υ
MS-5	-157.7336		
TS-3		-157.7042	337.15 i
MS-6	-157.7080		
TS-7b		-157.6596	2097.50 i
TS-7a		-157.6781	211.94 i
MS-7a	-157.7206		
MS-7b	-157.7483		

Table 4: Thermodynamic properties (ΔG and ΔG#) in kcal/mol and imaginary frequencies (υ) in cm^{-1} of the structures in the reaction path calculated at 298.15 K

The calculated structural parameters of the complexes are presented in Table 5. The average of the interatomic bonds Fe–Np in the complexes is of 2.01 Å, while the interatomic distance Fe-O for the two complexes TS-7b and MS-7b are 4.18 and 3.54 Å, respectively. A major changes was observed for the distance Fe-O in the acetone formation respect to the alcohol formation. This computed finding reflects probability/feasibility for the complex formation concerning the keto group during reaction. These *ab initio* geometries are in line with the preselected force field geometries with built-in atom types (Figure 2).

Structures	Fe-Np	Fe-O	∠ NpFeO
TS-3	2.03	1.92	105.0
MS-6	2.01	1.77	103.6
TS-7a	2.02	1.97	87.7
TS-7b	2.00	4.18	97.8
MS-7a	2.01	2.03	88.8
MS-7b	2.01	3.54	74.5

Table 5: Selected geometrical parameters of the minima and transition states optimized with B3LYP/6-31G plus the LanlDZ (bond lengths in Å and angles in degrees)

Discussion

Prior to docking, the identification of adequate geometries was required which were amenable to molecular mechanics formalisms dictated by predefined bond and atom types, their geometries, forces and charges without electron movements but electrostatics. Our proposed prerequisite might not always have been fulfilled in the literature (Table 2). Many authors do not even mention whether the reactive oxygen moiety was present or not, not to mention its attachment to the iron or substrate. In the present case of monooxygenase metabolism, the tetrahedral geometry of the substrate intermediate at the site of (mono)oxygenation was generated by an sp3 hybrid state to reflect the activated state of the aliphatic (C10-OH-PPC) or aromatic (6-OH-PPC, 7-OH-PPC, C2'-OH-PPC and C4'-OH-PPC) carbon atoms under catalytic attack by an oxygen moiety and iron IV-monooxygen radical cation.

Assumptions and model limitations

Molecular mechanics approach with predefined atom types and electron configurations – as well as quantum mechanics with elaborated and flexible electron descriptions – reflected the experimental data about substrate- or region-selectivities of selected CYP family members. As a silently made assumption, however, it has been assumed that the outer vestibule of CYP proteins with

the nearby surface of endoplasmic reticulum membranes would not contribute to regioselectivity. Another limitation was found in the warfarin-CYP2C9 crystal complex where the ligand was not directly bound at the catalytic site, but displaced. The enzyme does not show the wild type but a mutation type. Hence the displacement of ligand was ascribed to those mutated amino acids and the ligand position qualified as an artifact [45].

Conclusion

In the present docking study, Autodock poses were refined by Yeti and both were validated by selfdocking hitherto known liganded crystal complexes of enzymes. PPC and WFN as ligands were successfully docked into CYP2C9 and CYP3A4. Docking lend insight about the steric nature of the driving forces for substrate recognition. The docked poses simulated a ternary coordination complex (iron [IV] - activated oxygen - sp3 hybrid carbon atom). The docking results could reproduce the experimentally determined hydroxylation metabolites of PPC. The structures of four metabolites had already been identified by peaks in mass spectrometry (6-OH-PPC, 7-OH-PPC, C2'-OH-PPC and C4'-OH-PPC).The hitherto unknown structure of the last peak was computationally determined without ambiguity: it constitutes the PPC metabolite with hydroxylation at C10 position (acronym: 10-OH-PPC). Steric clashes for one of the five metabolites (C2'-OH-PPC) reflected induced fit which was simulated using an empiric rotamer library to find observed side chain conformations for Val113, Leu365 and Thr301. In addition, subtle geometrical and electronic movements were calculated by quantum mechanics to describe the reaction coordinate with the transition state geometry of the ligand-heme complex. Our results are all in line with reported regioselectivities. Although it would be farfetched to generalize on the theme, the technical achievement for the field of molecular modeling the enzymatic substrate metabolism becomes obvious: ligand - receptor docking algorithms (advanced force field calculations) can reproduce substrate binding to target enzyme despite the multiple step catalytic cycle.

Acknowledgement

Contribution ship

BK provided the experimental data about PPC metabolites, FMB provided the *ab initio* calculations of the transition states, KSA, IQM and TS carried out docking runs. TS created the study design, embedded the literature and prepared the Manuscript.

References

1. Zareh M, Davis A, Henderson S (2011) Reversal of warfarin-induced hemorrhage in the emergency department. West J Emerg Med 12: 386–92.

2. Reitsma PH, Van Der Heijden JF, Groot AP, Rosendaal FR, Büller HR (2005) A C1173T dimorphism in the VKORC1 gene determines coumarin sensitivity and bleeding risk. PLoS Med 2: e312.

3. Gonzalez FJ, Tukey RH (2006) Drug metabolism, de Goodman and Gilman's. The Pharmacological Basis of Therapeutics 11ed : 71-91.

4. He M, Korzekwa KR, Jones JP, Rettie AE, Trager WF (1999) Structural forms of phenprocoumon and warfarin that are metabolized at the active site of CYP2C9. Arch Biochem Biophys 372: 16-28.

5. Kirchheiner J, Ufer M, Walter EC, Kammerer B, Kahlich R, et al. (2004) Effects of CYP2C9 polymorphisms on the pharmacokinetics of R-and S-phenprocoumon in healthy volunteers. Pharmacogenetics 14: 19-26.

6. Kammerer B, Kahlich R, Ufer M, Schenkel A, Laufer S, et al. (2005) Stereospecific pharmacokinetic characterisation of phenprocoumon metabolites, and mass-spectrometric identification of two novel metabolites in human plasma and liver microsomes. Anal Bioanal Chem 383: 909-17.

7. Ufer M, Svensson JO, Krausz KW, Gelboin HV, Rane A, et al. (2004) Identification of cytochromes P 450 2C9 and 3A4 as the major catalysts of phenprocoumon hydroxylation in vitro. Eur J Clin Pharmacol 60: 173-82.

8. Ufer M, Kammerer B, Kahlich R, Kirchheiner J, Yasar Ü, et al. (2004) Genetic polymorphisms of cytochrome P450 2C9 causing reduced phenprocoumon (S)-7-hydroxylation in vitro and in vivo. Xenobiotica 34: 847-59.

9. Wester MR, Johnson EF, Marques-Soares C, Dijols S, Dansette PM, et al. (2003) Structure of mammalian cytochrome P450 2C5 complexed with diclofenac at 2.1 A resolution: evidence for an induced fit model of substrate binding. Biochemistry 42: 9335-45.

10. Seifert A, Pleiss. J (2009) Identification of selectivity-determining residues in cytochrome P450 monooxygenases: A systematic analysis of the substrate recognition site 5. Proteins 74: 1028-35.

11. Carlson HA (2002) Protein flexibility and drug design: How to hit a moving target. Curr Opin Chem Biol 6: 447-52.

12. Teodoro ML, Kavraki LE (2003) Conformational flexibility models for the receptor in structure based drug design. Curr Pharm Des 9: 1635-48.

13. Barril X, Fradera X (2006) Incorporating protein flexibility into docking and structure-based drug design. Expert Opin Drug Discovery 1: 335-49.

14. Sousa SF, Ribeiro AJ, Coimbra JT, Neves RP, Martins SA, et al. (2013) Protein-ligand docking in the new millennium–a retrospective of 10 years in the field. Curr Med Chem 20: 2296-314.

15. Zaretzki J, Bergeron C, Rydberg P, Huang TW, Bennett KP, et al. (2011) RS-Predictor: A new tool for predicting sites of cytochrome P450-mediated metabolism applied to CYP 3A4. J Chem Inf Model 51: 1667-89.

16. Kirchmair J, Williamson MJ, Tyzack JD, Tan L, Bond PJ, et al. (2012) Computational prediction of metabolism: sites, products, SAR, P450 enzyme dynamics, and mechanisms. J Chem Inf Model 52: 617-48.

17. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, et al. (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem 30: 2785-91.

18. Trott O, Olson A (2010) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31: 455-61.

19. Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) Development and validation of a genetic algorithm for flexible docking. J Mol Biol 267: 727-48.

20. Rarey M, Kramer B, Lengauer T, Klebe G (1996) A fast flexible docking method using an incremental construction algorithm. J Mol Biol 261: 470-89.

21. Alonso H, Bliznyuk AA, Gready JE (2006) Combining docking and molecular dynamic simulations in drug design. Med Res Rev 26: 531-68.

22. Nabuurs SB, Wagener M, Vlieg JD (2007) A flexible approach to induced fit docking,» Journal of medicinal chemistry. 50: 6507-18.

23. Tantar AA, Conilleau S, N, Parent B, Melab Brillet L, et al. (2008) Docking and biomolecular simulations on computer grids: Status and trends. Curr Comput Aided Drug Des, 4: 235-49.

24. Vasanthanathan P, Lastdrager J, Oostenbrink C, Commandeur J N, Vermeulen NP, et al. (2010) Identification of CYP1A2 ligands by structure-based and ligand-based virtual screening. Med Chem Comm, 2: 853-59.

25. Lo SN, Chang YP, Tsai KC, Chang CY, Wu TS, et al. (2013) Inhibition of CYP1 by berberine, palmatine, and jatrorrhizine: selectivity, kinetic characterization, and molecular modeling. Toxicol Appl Pharmacol 272: 671-80.

26. Salminen KA, Rahnasto-Rilla M, Väänänen R, Imming P, Meyer A, et al. Time-Dependent Inhibition of CYP2C19 by Isoquinoline Alkaloids: In Vitro and In Silico Analysis. Drug Metab Dispos 43: 1891-904.

27. Kumalo HM, Bhakat S, Soliman ME (2015) Theory and applications of covalent docking in drug discovery: merits and pitfalls. Molecules 20: 1984-2000.

28. Kato H, Yamaotsu N, Iwazaki N, Okamura S, Kume T, et al. (2017) Precise prediction of activators for the human constitutive androstane receptor using structure-based three-dimensional quantitative structure–activity relationship methods. Drug Metab Pharmacokinet 32: 179-88.

29. Vedani A, Huhta DW (1991) Algorithm for the systematic solvation of proteins based on the directionality of hydrogen bonds. J Ame Chem 113: 5860-2.

30. Gaussian I, (2009) Gaussian 9 Revission A 01. Wallingford CT.

31. Tripos, Sybyl X (2010) St Louis MO 63144-2319USA: Tripos Inc.

32. Bernstein FC, Koetzle TF, Williams GJ, Meyer EF, Brice MD (1977) The protein data bank. A computer-based archival file for macromolecular structures. Eur J Biochem 80: 319-24.

33. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling,» Electrophoresis 18: 2714-23.

34. Pedretti A, Villa L, Vistoli G (2002) VEGA: a versatile program to convert, handle and visualize molecular structure on Windows-based PCs. J Mol Graph Model 21: 47-9.

35. M Clark, RD. Richard DC, Nicole Van O (1989) Validation of the general purpose Tripos 5.2 force field. J Computational Chem 10: 982-1012.

36. Vedani A, Huhta DW (1990) A new force field for modeling metalloproteins. J Am Chem Society 112: 4759-67.

37. Yano JK, Hsu MH, Griffin KJ, Stout CD, Johnson EF (2005) Structures of human microsomal cytochrome P450 2A6 complexed with coumarin and methoxsalen. Nat Struct Mol Biol 12: 822-3.

38. Wester MR, Yano JK, Schoch GA, Yang C, Griffin KJ, et al. (2004) The structure of human cytochrome P450 2C9 complexed with flurbiprofen at 2.0-Å resolution. J Biol Chem 279: 35630-7.

39. Guroff G, Daly JW, Jerina DM, Renson J, Witkop B, et al. (1967) Hydroxylation-induced migration: the NIH shift. Recent experiments reveal an unexpected and general result of enzymatic hydroxylation of aromatic compounds. Sci 157: 1524-30.

40. Mancy A, Broto P, Dijols S, Dansette PM, Mansuy D (1995) The substrate binding site of human liver cytochrome P450 2C9: an approach using designed tienilic acid derivatives and molecular modeling,» Biochem 34: 10365-75.

41. Poulos TL, Finzel BC, Howard AJ (1987) High-resolution crystal structure of cytochrome P450cam. J Mol Biol 195: 87-700.

42. Raag R, Poulos TL (1989) Crystal structure of the carbon monoxide-substrate-cytochrome P-450CAM ternary complex. Biochem 28: 7586-92.

43. Batabyal D, Poulos TL (2013) Crystal structures and functional characterization of wild-type CYP101D1 and its active site mutants. Biochem 52: 8898-906.

44. Williams PA, Cosme J, Vinkovic DM, Ward A, Angove HC, et al. (2004) Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone. Science 305: 683-6.

45. Williams PA, Cosme J, Ward A, Angove HC, Matak Vinković D, et al. (2003) Crystal structure of human cytochrome P450 2C9 with bound warfarin. Nature 424: 464-8.

46. Mo SL, Zhou ZW, Yang LP, Wei MQ, Zhou SF (2009) New insights into the structural features and functional relevance of human cytochrome P450 2C9. Curr Drug Metab10: 1075-126.

47. Zhou SF, Zhou ZW, Yang LP, Cai JP (2009) Substrates, inducers, inhibitors and structure-activity relationships of human Cytochrome P450 2C9 and implications in drug development. Curr Med Chem 16: 3480-675.

48. Belic A, Temesvári M, Kohalmy K, Vrzal R, Dvorak Z, et al. (2009) Investigation of the CYP2C9 induction profile in human hepatocytes by combining experimental and modelling approaches. Curr Drug Metab 10: 1066-74.

49. Shapovalov MV, Dunbrack RL (2011) A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. Structure 19: 844-58.

50. Flora DR, Rettie AE, Brundage RC, Tracy TS (2017) CYP2C9 Genotype-Dependent Warfarin Pharmacokinetics: Impact of CYP2C9 Genotype on R-and S-Warfarin and Their Oxidative Metabolites. J Clin Pharmacol 57: 382-93.

51. Jones DR, Kim SY, Guderyon M, Yun CH, Moran JH (2010) Hydroxywarfarin metabolites potently inhibit CYP2C9 metabolism of S-warfarin. Chem Res Toxicol 23: 939.

52. A Becke (1993) Density-functional thermochemistry. III. The role of exact exchange. J Chem Phys 98: 5648-52.

53. Zhao Y, D Truhlar (2008) Density Functionals with Broad Applicability in Chemistry. Acc Chem Res 41: 157-67.

54. Walker M, Harvey AJ, Sen A, Dessent CE (2013) Performance of M06, M06-2X, and M06-HF Density Functionals for Conformationally Flexible Anionic Clusters: M06 Functionals Perform Better than B3LYP for a Model System with Dispersion and Ionic Hydrogen-Bonding Interactions. J Phys Chem A 117: 12590-600.

55. Hariharan PC, Pople JA (1972) The effect of d-functions on molecular orbital energies for hydrocarbons. Chem Phys Lett 16: 217-9.

56. Hay PJ, Wadt RW (1985) Ab initio effective core potentials for molecular calculations. Potentials for K to Au including the outermost core orbitals. J Chem Phys 82: 299-310.

57. R. Dennington T. Keith, M Millam, S Inc y K Shawnee Mission (2016) GaussView, Version 5.0.8.

58. JP. Collman, JU Hoard, N Kim, G Lang, CA Reed (1975) Synthesis, stereochemistry, and structure-related properties of alpha, beta, gamma ,delta-tetraphenyl-porphin-atoiron(II). J Am Chem Soc 97: 2676-8.

59. H Li, S Narasimhulu, LM Havran, JD Winkler, TL. Poulos (1995) Crystal structure of cytochrome P450cam complexed with its catalytic product, 5-exo-hydroxycamphor. J Am Chem Soc 117: 6297-9.

60. Schmidt TC, Welker A, Rieger M, Sahu PK, Sotriffer CA, et al. (2014) Protocol for rational design of covalently interacting inhibitors. Chemphyschem. 15: 3226-35.

61. Du H, Li J, Cai Y, Zhang H, G Liu, Tang Y, Li W (2017) Computational Investigation of Ligand Binding to the Peripheral Site in CYP3A4: Conformational Dynamics and Inhibitor Discovery. J Chem Inf Model 57: 616-26.