Combined Docking and Quantum Chemical Study on CYP-mediated Metabolism of Estrogens in Man

Anikó Lábas[1,#], *Balázs Krámos*[1,2,#], *Julianna Oláh*[1,*]

[1] Department of Inorganic and Analytical Chemistry, Budapest University of Technology and Economics, Szent Gellért tér 4, H-1111 Budapest, Hungary

[2] Institute of Organic Chemistry Research Centre for Natural Sciences, Hungarian Academy of Sciences, Magyar Tudósok körútja 2, Budapest, P.O. Box 286, 1519, Hungary

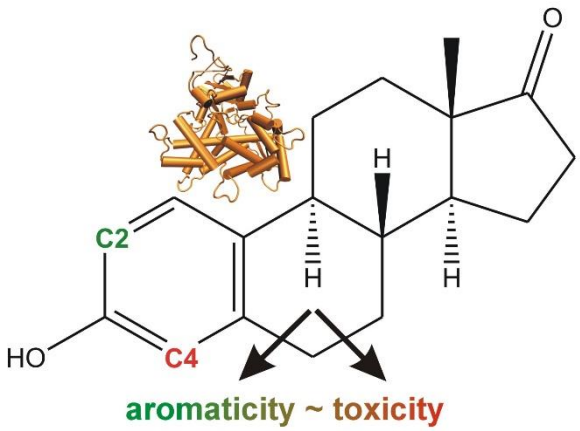
Keywords cytochrome P450 enzymes, modelling, quantum chemistry, docking, estrone, equilin, equilenin

---

[#] these two authors contributed equally to this work

[*]corresponding author, to whom corresponding should be addressed.

email: julianna.olah@mail.bme.hu



TOC graphic

Abstract

Long-term exposure to estrogens seriously increases the incidence of various diseases including breast cancer. Experimental studies indicate that cytochrome P450 (CYP) enzymes catalyze the bioactivation of estrogens to catechols, which can exert their harmful effects via varies routes. It has been shown that the 4-hydroxylation pathway of estrogens is the most malign, while 2-hydroxylation is considered a benign pathway. It is also known experimentally that with increasing unsaturation of ring B of estrogens the prevalence of the 4-hydroxylation pathway significantly increases. In this study we used a combination of structural analysis, docking and quantum chemical calculations at the B3LYP/6-311+G* level to investigate the factors that influence the regioselectivity of estrogen metabolism in man. We studied the structure of human - estrogen metabolizing enzymes (CYP1A1, CYP1A2, CYP1B1, CYP3A4) in complex with estrone using docking, and investigated the susceptibility of estrone, equilin and equilenin (which only differ in the unsaturation of ring B) to undergo 2- and 4-hydroxylation using several models of CYP enzymes (Compound I, methoxy and phenoxy radical). We found that even the simplest models could account for the experimental difference between the 2- and 4- hydroxylation pathways, thus might be used for fast screening purposes. We also show that reactivity indices, specifically in this case the radical and nucleophilic condensed Fukui functions also correctly predict the likeliness of estrogen derivatives to undergo 2- or 4-hydroxylation.

1. Introduction

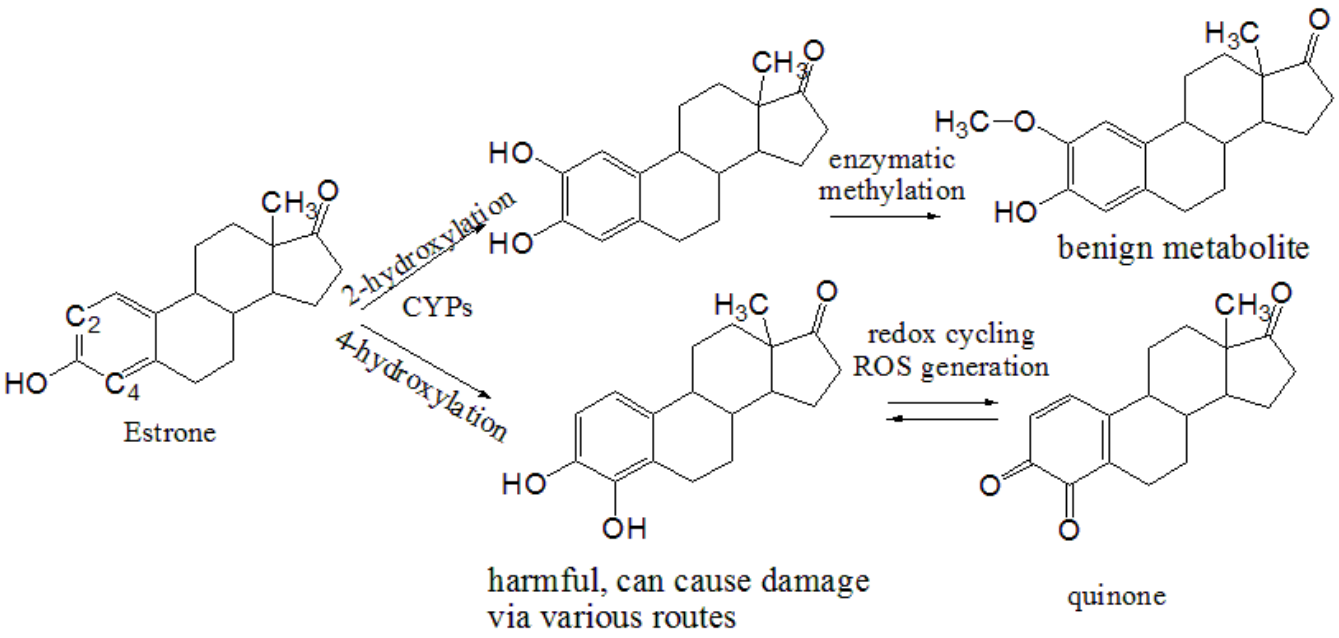
Cytochrome P450 enzymes (CYP) are among the most versatile biocatalyst known today,[1] besides α-ketoglutarate-dependent enzymes. Their common feature is a cysteinate-ligated heme group, which via an elaborate catalytic cycle activates molecular oxygen. Most CYPs function as monooxygenases, and catalyze a variety of reactions, including aromatic and aliphatic hydroxylations, epoxidation, dealkylations and heteroatom oxidations. CYPs are the most important detoxificating enzymes in the human body, they contribute to the metabolism of about 80-90% of the drugs. Although, in most cases CYP-mediated metabolism leads to the increased solubility of the xenobiotics and thereby facilitates their secretion and reduces their toxicity, in other cases compounds could be bioactivated to harmful substances; this latter aspect is frequently exploited in the case of e.g. anti-cancer agents.[2]

In the last few decades, more and more information has been gathered that CYP enzymes may also bioactivate natural substances, such as hormons to toxic agents. The most striking example is the case of estrogens. Estrogens are essential and multifunctional steroid hormones in the human body. Most important of all, they participate in the estrous cycle of women, but also influence arterial vasodilation, the maintenance of bone density, and aid in neuroprotective actions. Long-term exposure to estrogens either via early menarche or late menopause or hormon replacement therapies (HRT) has been shown to be strongly linked to increased risk of breast cancer.[3–7] HRTs were developed in the 1940s to alleviate the symptoms of menopause, e.g. hot flashes and night sweats, and these treatments have also helped in the prevention of bone loss due to low levels of estrogens. Since the introduction of HRTs to the market, the drug development and approval processes have become much more complex requiring, among others, to test the long-term effects of drug candidates. In the last two decades similar large-scale studies were carried out for HRTs including the Women's Health Initiative in the US and the Million Women's study in the UK,[8] both of which revealed an increased risk in breast cancer due to the use of HRT.

Despite the clear indication that estrogens exert carcinogenic activity, the underlying mechanisms are still ill-understood[9–12] partly due to the fact that cancer usually arises as a consequence of a series of mutations.[13] At least three different pathways have been suggested to be involved in steroidal estrogen

carcinogenesis. Estrogens may exert their harmful effect via a hormonal pathway, by which estrogen stimulates cell proliferation through nuclear endoplasmic reticulum-mediated signaling pathways leading to an increased risk of genomic mutations during DNA replication.[12,14] Membrane-associated endoplasmic reticuli, which regulate extranuclear estrogen signaling, may also contribute to carcinogenesis.[15,16] The third pathway involves the redox active species generated in CYP-mediated metabolism of estrogens,[17,18] including estrogen catechols, which can be easily oxidized to estrogen *o*-quinones. Redox cycling of cathecols, quinones and semiquinones leads to the generation of reactive oxygen species which may modify proteins and nucleic acids, leading to mutations and initiate carcinogenesis.[19]

Various metabolic routes of estrogens are known in the human body. At least four CYP isoforms (CYP1A1, CYP1A2, CYP1B1 and CYP3A4) have been suggested to metabolize estrogens leading to various products. The chemical identity of products formed by CYPs are influenced by several factors, including (1) the active site structure of the enzymes ultimately determining the binding affinity and orientation of substrates in the active site, (2) isoform abundance and distribution in various tissues, (3) reactivity of functional groups of the substrate towards oxidation, (4) CYP induction effects (various compounds may influence the transcription factors that are important for the expression of CYP isoforms,[20] and not uniformly elevated CYP levels could lead to changed product ratios (e.g. if two different isoforms metabolize the same compound leading to different products). In the case of estrogens, the most prevalent metabolic routes are 2- and 4-hydroxylation (see Scheme 1.), and to a lesser extent 16-hydroxylation, which is not considered herein as it involves a completely different part of the molecule.



Scheme 1. 2- and 4-hydroxylation pathways of estrogens catalyzed by cytochrome P450 enzymes represented through the example of estrone. The produced 3,4-cathecols easily autooxidize to quinones and redox cycling between catechols and quinones leads to the production of reactive oxygen species (ROS).

2-hydroxylation is considered a benign pathways as the produced 2-hydroxy estrogens[21,22] are rapidly methylated by cathecol-o-methyl transferase producing 2-methoxy estrogens, which are known to be non-carcinogenic and to inhibit the proliferation of cancer cells. In contrast, 4-hydroxylation yields 3,4-cathecols which can be easily oxidized by peroxidases to produce 3,4-semiquinones and 3,4-quinones. Recently, estradiol-3,4-quinone was identified as a potential risk factor in breast cancer,[23] and it was shown to form a quinone-DNA adduct thereby causing mutations and initiating carcinogenesis. Furthermore, Liehr and Ricci reported that 4-hydroxylation of estradiol is a marker for human mammary tumors.[24] In the mammary gland CYP1A1 and CYP1B1 are considered to be the major estrogen-metabolizing enzymes. Using homology modeling and docking Yamamoto et al suggested that different orientation of estradiol within the two active sites is responsible for the differing regioselectivity of these isoforms.[25] However, the active site structure of enzymes is not the only factor for determining the regioselectivity of the enzyme. Intrinsic reactivity differences between the sites of the substrate may also

contribute to the obtained product ratios.[26,27] This may especially be the case for the CYP3A4 isoform, which is known to possess a large, non-polar cavity which may simultaneously accommodate even two large substrates. In this case the most reactive site of the substrate might be the most likely to react. Recently, Dai et al set up a mechanistic model to predict the metabolic routes of various steroids.[28]

In the US the most widely used form of HRT is Premarin (Wyeth-Ayerst), which contains at least 10 different estrogen derivatives including estrone, equilin and equilenin (see Chart 1), which differ only in the unsaturation of Ring B.

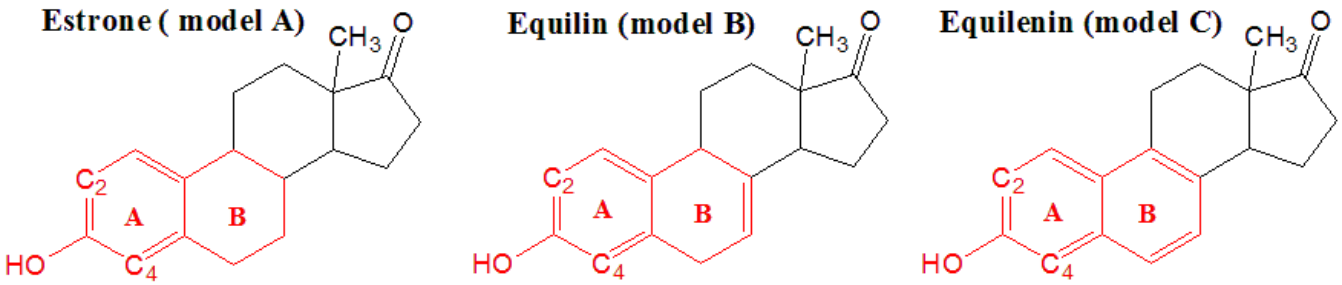


Chart 1. Structures of estrone, equilin, equilenin with experimentally observed 2- vs 4-hydroxylated product ratios produced by baboon liver microsomes.[29] Model compounds used for the quantum mechanical modelling calculations are shown in red: **A** model of estrone:, **B** model of equilin, **C** model of equilenin.

Using baboon liver microsomes it was shown that the ratio of 2- and 4-hydroxylation pathways is strongly dependent (see data in Chart 1) on ring B of the estrogen derivatives and that from primarily 2-hydroxylation of estrone it turns to exclusively 4-hydroxylation for equilenin.[29] Similarly, for 2-naphtol, which contains the same aromatic system as equilenin, only 4-hydroxylation occurs when studied with hamster liver microsomes.[30] As the CYP3A4 enzyme is the most abundant CYP present in the liver and the structure of these estrogens is very similar, these experimental findings may indicate that in the case

of these experiments the intrinsic reactivity difference between positions C_2 and C_4 of estrogens significantly influences the obtained product ratios. This question is especially important because when CYP3A4 staining was assessed semi quantitatively the mean staining score values of CYP3A4 were significantly higher in tumors compared to that of normal breast tissues implying that CYP3A4 is expressed in both tumor and normal breast tissue with an increased expression in tumours.[31] From this it follows that in tumors there is a larger quantity of CYP3A4 present, which may produce larger amounts of 4-hydroxylated products further worsening the situation in breast cancer tissues.

For this reason, it is important to understand the factors that influence the 2- and 4-hydroxylation pathways of estrogens. A possible way, among others, to get insight into these factors is to use quantum chemical modeling, which allows one to determine the energy, structure, charge distribution and other relevant properties of species (i.e. reactants, products, transition states and intermediates) involved in the metabolic reaction, thereby giving insight at the atomic level into mechanism of the reaction.

The chemistry of CYP-enzyme-catalyzed reactions has been studied in detail,[32,33] including the properties of heme species involved in the catalytic cycle of CYPs, mechanism of alkane hydroxylation, mechanism of hydride transfer in $P450_{nor}$ enzyme,[34] the biosynthesis of estrogens,[35,36] and the electronic structure of the oxidizing species (Compound I) of CYPs. In a recent combined quantum chemical molecular mechanics study (QM/MM) we showed that Compound I seems to exhibit very similar properties in different CYP enzymes, implying that most likely differences in Compound I do not significantly contribute to the regioselectivity differences observed for CYPs.[35]

In the present work we set out to study using a diverse set of modelling tools the factors that influence the 2- vs 4-hydroxylation pathways of estrogen metabolism. In the first part of the paper using docking and structural considerations we investigate the effect of protein active site architecture on the likelihood of 2- vs 4-hydroxylations by CYP1A1, CYP1B1 and CYP3A4, the isoforms contributing most significantly to estrogen metabolism by man. In the second part of our work we use quantum mechanical (QM) calculations to study the electronic factors that affect the 2- and 4- hydroxylation pathways We would like to note that the QM calculations that are presented herein reveal the tendency of Compound I

to oxidize C_2 or C_4 in estrogen derivatives, however, further role of the protein environment (e.g. orientation of the ligand in the pocket, possible polarizing effect of the enzyme) was not taken into account in these calculations.

2. Computational details

2.1. Molecular modeling and docking.

In order to have insight into the similarity of the amino acid sequences of the estrogen metabolizing enzymes their sequences were aligned using Clustal Omega,[38–40] an online bioinformatics tool. Afterwards, we searched for co-crystalized X-Ray structures of the estrone-CYP (CYP1A1, CYP1A2, CYP1B1 and CYP3A4) complexes in order to get information on the binding mode of estrogens in the various active sites. However, these are not available, therefore molecular docking methods were used to generate the protein-ligand complexes. The amino acid side chains in the substrate free X-ray structures of various isoforms prevented the docking of the estrone ligand into the close proximity of the heme group in the active site (e.g. Tyr212 side chain is in close contact with the heme group in CYP3A4 enzyme variant leaving no space for the substrate), therefore proteins in complex with inhibitors were used (PDB IDs: CYP1B1–3PM0,[41] CYP1A1–4I8V,[42] CYP1A2-2HI4,[43] CYP3A4–4I4G[44]). After the elimination of the inhibitors from the structures, the missing residues inside the sequences were replaced using MODELLER 9.14.[45,46] Structures were prepared for docking in the following way. Crystal waters were eliminated from all systems, as these were all found far from the active site, and were unlikely to effect the results of the docking studies. Only polar hydrogen atoms were added to the structures using the built-in function of the docking program. The protonation state of the ionizable residues of the enzymes was predicted using the PROPKA 3.1 program package,[47–50] and all ionizable were predicted to be present in their standard protonation state, with the exception of Asp182 in CYP3A4 (4I4G) and Glu466 in CYP1A1 (4I8V). As none of these two residues lie reasonably in the active site, their protonation state in unlikely to change the predicted orientation of the ligand in the active site. For this reason, all ionizable residues were protonated according to their standard protonation state.

The structure of the estrone ligand was taken from the DrugBank database (Accession Number: DB00655),[51] and only a single hydrogen atom was added to the structure, the hydrogen atom of the hydroxyl group in ring A, as all other hydrogen atoms are non-polar hydrogen atoms in estrone. The AutoDock Vina[52] docking program was used to generate the enzyme-ligand complexes for each isoform. For ranking the docked structures, the own scoring function of the program was used, which extracts empirical information from both the conformational preferences of the receptor-ligand complexes and the experimental affinity measurements.[52] In the case of the CYP1B1 isoform manual docking was also done.

The docking methodology was validated by two different methodologies. (1) The inhibitors that were co-crystalized with the enzymes (desoxyritonavir analogue for CYP3A4; alpha-naphthoflavone for CYP1A1, CYP1A2 and CYP1B1) were docked back into the active sites of the enzymes using the same methodology. For members of the CYP1 family, very good agreement was obtained with the experiment, although the docking modes closely ressembling the X-ray structure were predicted to have slightly lower affinities than the highest-affinity structure (see Table S1 in the supporting information). In the case of CYP3A4 the docked structures were different from the original pose due to the very large cavity and to the fact that estrone has very few polar groups and the active site of CYP3A4 is also mainly borderd by non-polar groups thus there is little opportunity for strong specific interactions that would determine the position of the ligand in the active site. (2) Estrone was docked into several X-Ray structures of the same enzyme variant in order to check the dependence of the obtained results on the used X-Ray structure. Unfortunately, for members of the CYP1 family we only found one X-Ray structure, however, in the case of CYP3A4 many different X-Ray structures have been solved that we used for docking: PDB ids: 4NY4,[53] 4K9X,[54] 4K9W,[54] 4K9V,[54] 4K9U,[54] 4K9T,[54] 4I4H,[44] 4D7D,[55] 4D78,[55] 4D75,[55] 4D6Z,[55] 3UA1,[56] 3TJS,[57] 3NXU,[58] 2V0M,[58] 2J0D[59]). From the obtained results we concluded that the docking methodology is likely to yield relevant docking poses for our systems, especially for CYP1 family members.

2.2 Quantum mechanical (QM) calculations.

We carried out QM calculations on model systems of estrone, equilin and equilenin (see the atoms colored in red in Chart 1. Model of estrone will be called **A**, model of equilin will be called **B**, and model of equilenin will be called **C**. We modeled the protoporphyrin IX ring present in CYP enzymes as an unsubstituted porphine ring (see Chart 2.), as it was earlier shown to be a good model of it.[60] The axial cysteine moiety was modeled as SH^- group. We also studied small model compounds of Compound I, the methoxy and phenoxy radicals in order to investigate the possibility to reduce the computational cost (Chart 2.). The calculations were carried out with the Gaussian 09 program package[61] using the B3LYP functional with the 6-31G(d) basis set on all atoms, except on iron, on which the SDD pseudo potential has been used in conjunction with the corresponding basis set, as was done earlier.[34–36] This basis set composition will be called BS1. In order to obtain reliable energies, single point calculations were performed on all structures at the B3LYP/6-311+G* level of theory. Second derivative calculations have been carried out at the B3LYP/BS1 level of theory to verify the identity of all located stationary points, transition structures were characterized by a single imaginary frequency and minima by only positive frequencies. Zero-point vibration energy (ZPE) correction to the energies is included at the B3LYP/BS1 level of theory. Dispersion correction was calculated using the D3 method[62] developed by Grimme with Becke-Johnson damping.[63]

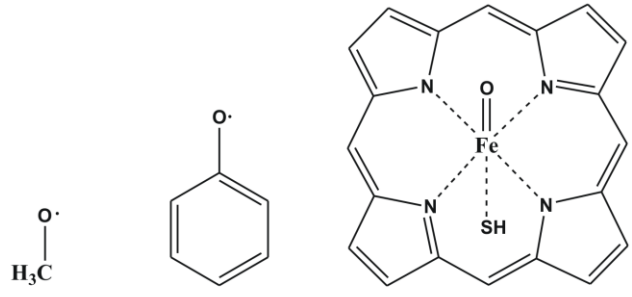


Chart 2. Models of Compound I, the active form of CYP enzymes used in this study. Non-polar hydrogen atoms from the phenoxy radical and Compound I were omitted for clarity.

All calculations were carried out in the doublet state, as it was trivial for the methoxy and phenoxy radical calculations. In the case of compound I the reactions may happen both in the quartet or the doublet states, but earlier calculations clearly show that aromatic carbon oxidation has a lower barrier on the doublet surface.[64,65] The doublet state calculations were preceded by calculations in the quartet state and the molecular orbitals were checked to ensure that the correct doublet state was obtained, where the spin of the electron found on the molecular orbital composed of a combination of a porphyrin π orbital with idealized a_{2u} symmetry and a p orbital on sulfur was opposite to the spins of the electrons found in the molecular orbitals involving the Fe-O bond. The singly occupied molecular orbitals of compound I are shown in Figure 1.

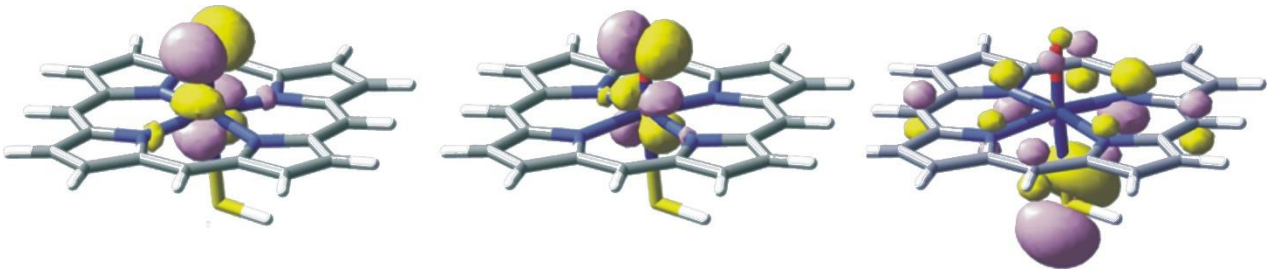


Figure 1. Singly occupied molecular orbitals of Compound I.

Reactivity indices[66] have been calculated based on eq (1), eq (2) and eq (3) at the B3LYP/6-311+G* level of theory from the NPA charges[67] of the neutral, positively and negatively charged species, all calculated at the optimized geometry of the neutral species: Fukui function for a nucleophilic attack (f_A^+) provoking an electron increase in the system, Fukui function for electrophilic attack (f_A^-) provoking an electron decrease in the system, and (f_A^0) for a radical attack on the system.[68]

$$f_A^+ = q_{A,N_0+1} - q_{A,N_0} \qquad (1)$$

$$f_A^- = q_{A,N_0} - q_{A,N_0-1} \qquad (2)$$

$$f_A^0 = \frac{f_A^+ + f_A^-}{2} \qquad (3)$$

3. Results and Discussion

First, we will compare the sequence and structure of estrogen-metabolizing CYP enzymes and study the binding mode of estrone in their active sites. This will be followed by the description of the electronic factors that influence the 2- vs 4-hydroxylation pathways. Naturally, the results of quantum chemical calculations should only be cautiously compared to the experimentally observed product ratios, as experimental selectivities are determined by various factors including isoform abundance and their regioselectivities. Furthermore liver microsome preparations may contain a variety of CYP isoforms, which was not determined in the experiment.[29,30] In any case we believe that the results of our quantum mechanical calculations do shed light on the intrinsic reactivity of C_2 and C_4 of estrogens and in the case of isoforms with large, non-polar active sites (i.e. CYP3A4) this may be the major determining factor of the preferred hydroxylation routes. Finally, the possibility of using smaller model compounds of Compound I in order to study the intrinsic reactivity of C_2 and C_4 of estrogens will be investigated and it will be checked whether reactivity indices (which solely require a simple quantum chemical calculation on the substrate) can be used to explain the observed reactivity of C_2 and C_4 in the studied molecules.

3.1. Molecular modelling and docking

The human 3A4, 1A1, 1B1 and 1A2 isoforms of CYP enzymes were investigated. The aligned protein sequences are shown in Figure 2 and the sequence identity matrix in Table 1. Protein sequence alignment is a useful way of arranging known protein sequences to identify amino acid residues or regions in the sequences that remained conserved during evolution. Residues that are identical or similar in chemical behavior can be characterized between sequences which can enable us to identify the important structural and functional protein elements between various isoforms. Gaps are frequently inserted between residues in order to reach the best alignment. In order to quantify the identity between the sequences the Percent Identity Matrix was calculated for each pairwise sequence showing the extent to which two sequences are invariant.[69] The 1A1 and 1A2 isoforms show relatively high sequence identity (over 70%), as it is expected for isoforms belonging to the same subfamily of CYP enzymes (see letter A in their name). Such a large identity between two enzymes, strongly suggest that they have a very similar 3D structure. Protein

1B1 shows about 40% identity with both 1A1 and 1A2 in accordance with the fact all three of them have been classified within the CYP1 family. The sequence of 3A4 is very different from the other ones, it shows only 20% identity to them.

Table 1. Percent Identity Matrix created by Clustal 2.1 for the selected four CYP isoforms.[38–40]

	3A4	**1B1**	**1A1**	**1A2**
3A4	100.0	21.0	22.3	22.6
1B1	21.0	100.0	40.8	38.7
1A1	22.3	40.8	100.0	72.2
1A2	22.6	38.7	72.2	100.0

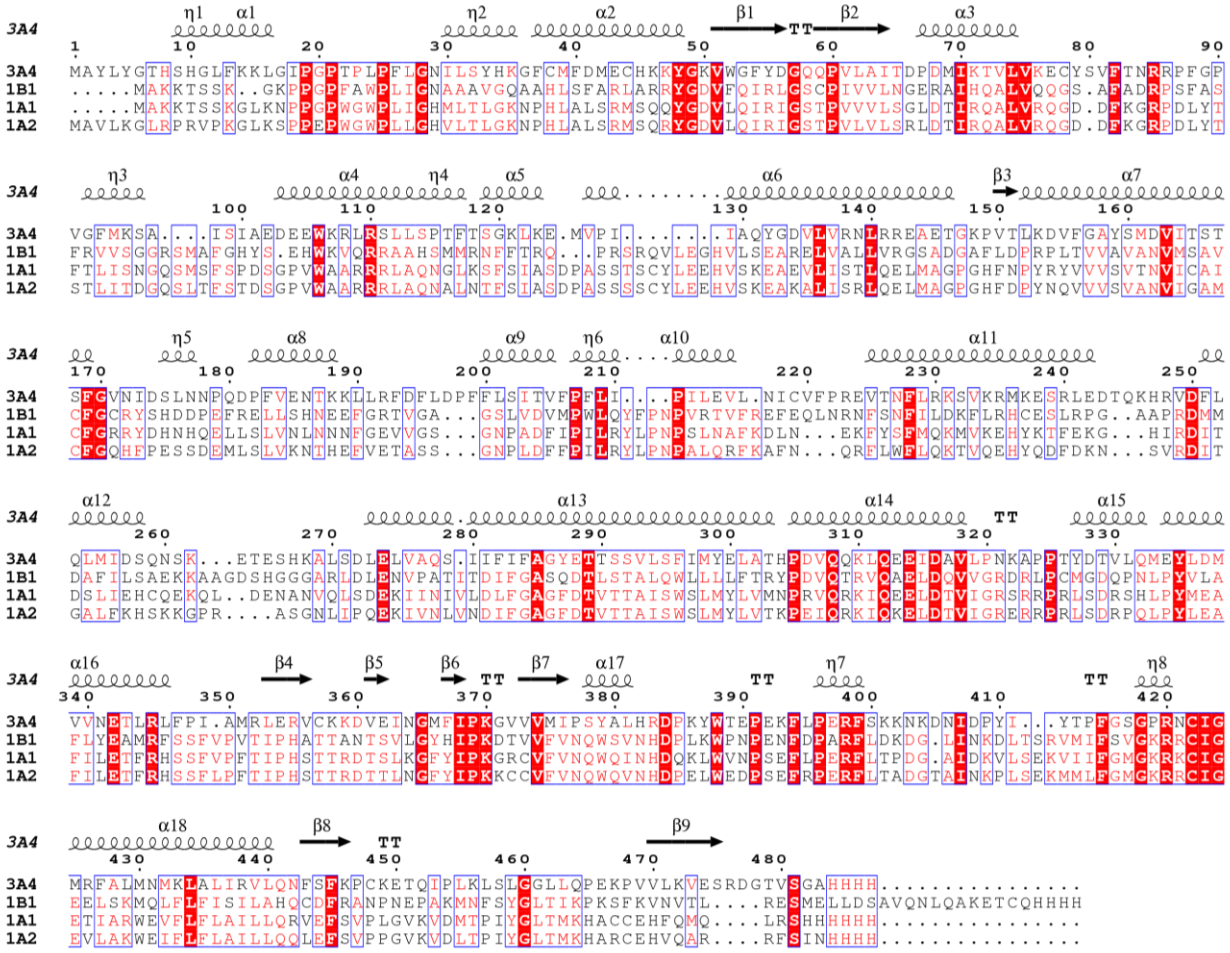


Figure 2. Aligned protein sequences of CYP 3A4, 1B1, 1A1 and 1A2 CYP isoforms**.** The numbering and the secondary structure elements are according to CYP3A4, using squiggles (α-helices), arrows (β-strands), lines (turns), TT (strict β-turns) and TTT (strict α-turns). The layout was generated by ESPript 3.0.[70] In the alignment, red boxes with white characters indicate strict identity and red characters show similarity, respectively.

Despite the varying degrees among the sequence identities of the four isoforms, the three dimensional structure of the substrate binding sites above the heme group shows a certain similarity in all studied isoforms: mainly α-helices and loops surround it. The most significant differences among the active sites of the members of the CYP1 family and that of CYP3A4 is the presence of an α-helix parallel with the heme group at the top of the binding site in the members of the family 1, leading to a significantly narrowed binding space (Figure 3) compared to CYP3A4. This is very well demonstrated by differences in volume of the binding cavities in the four isoforms. The POCASA 1.1 webserver[71] (with its standard set up protocol) predicted the following cavity volumes: CYP3A4: average volume for the studied 16 X-rays structures: 1212 ± 489 $Å^3$, and for PDB ID 4I4G: 1043 $Å^3$, CYP1A1: 332 $Å^3$, CYP1A2: 377 $Å^3$ and CYP1B1: 384 $Å^3$. This difference is the most likely explanation for the much wider substrate specificity of CYP3A4 enzymes compared to CYP1 enzymes.

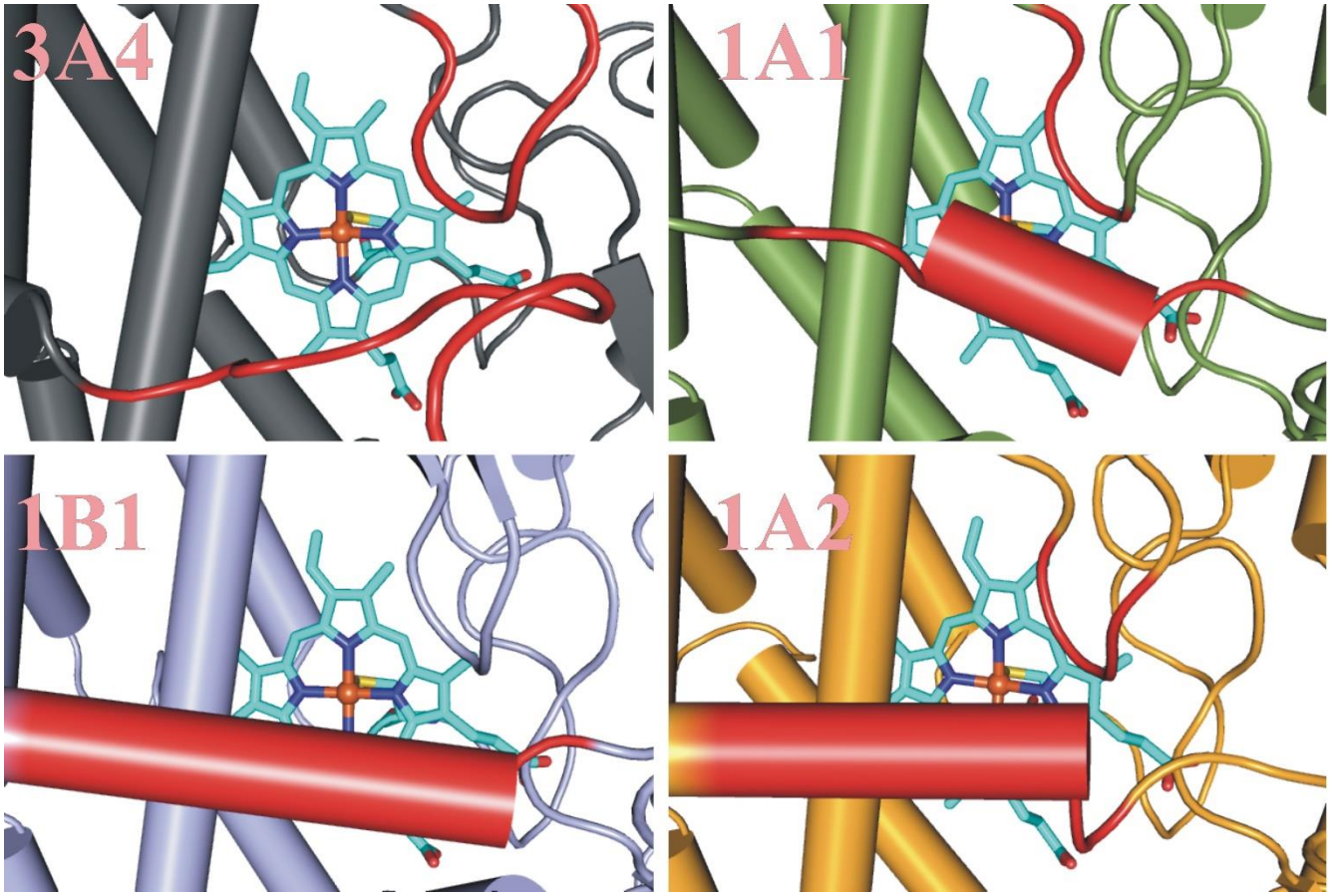


Figure 3. Ligand binding site of the investigated CYP enzyme isoforms. α-helices are represented by cartoons and loops by tubes. Structural elements colored in red reduce the size of the binding site of CYP1 enzymes. Color coding of the heme group: light blue: carbon, dark blue: nitrogen, iron: red sphere.

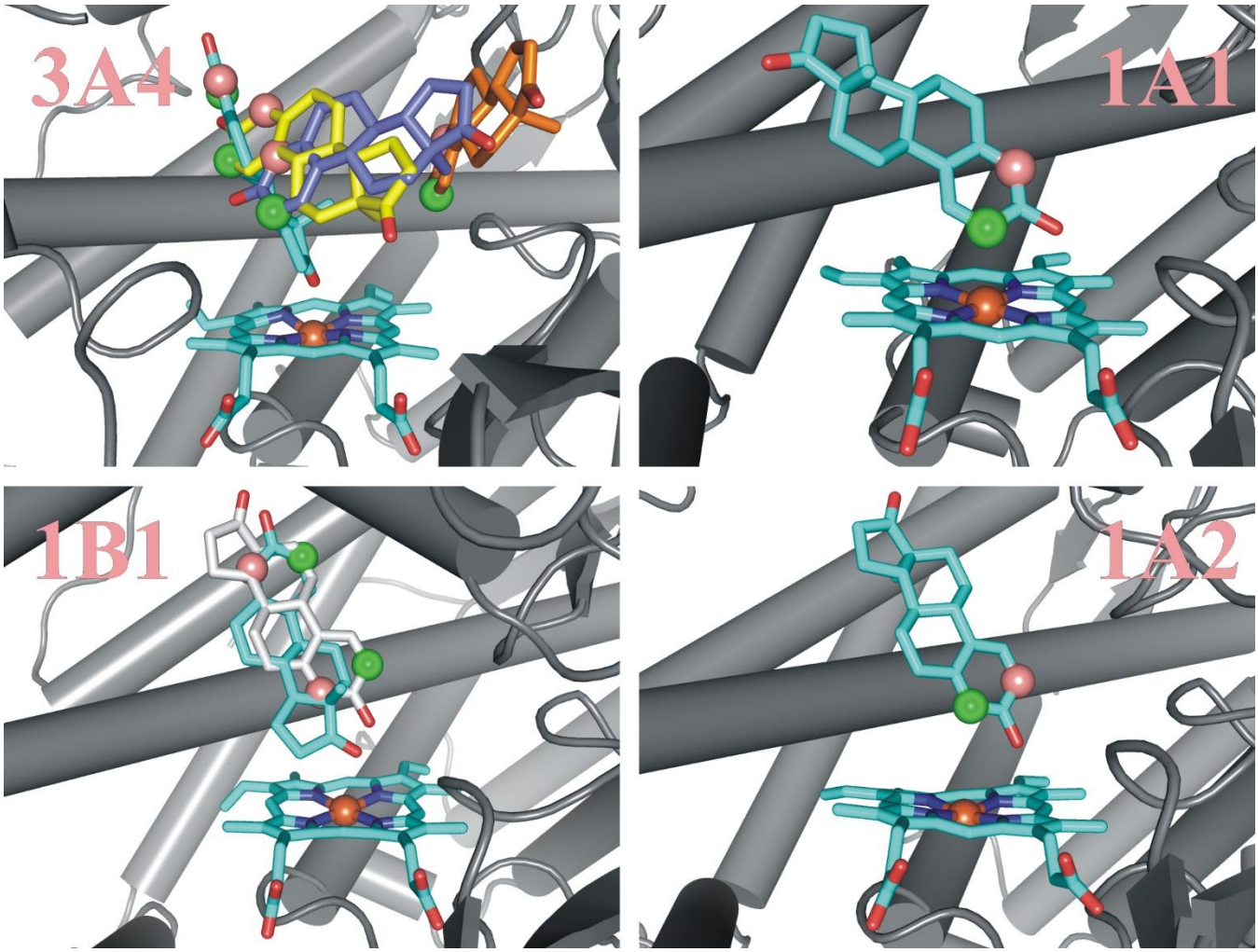


Figure 4. Representative binding modes of estrone produced by AutoDock Vina. α-helices are represented by cartoons and loops by tubes. In CYP3A4 the possible ligand orientations are shown using different colors for its representation. In CYP1B1 rotated binding mode for the appropriate hydroxylation is shown in grey. The C_2 carbon is indicated by a salmon-coloured sphere and C_4 by a green sphere in all isoforms. Color coding: red: oxygen, light blue: carbon, dark blue: nitrogen, iron: red sphere.

After having examined the active sites of the four CYP isoforms we docked estrone into their active sites using AutoDock Vina.[52] Only estrone was docked, as equilin and equilenin have a very similar 3D structure to estrone, thus we expect them to have very similar binding modes in the active sites. The docking yielded various poses for estrone in each isoform, that are depicted in Figures S1-S4 in the SI, and the binding affinities are shown in Table S2. In the case of CYP3A4 large flexibility of the poses is

observed, which is due to the very spacious active site (representative examples are shown in Figure 4). Most of these structures lie quite far from the heme group as the ligand interacts with the non-polar groups bordering the active site and an empty space is found between the heme group and the ligand, thus in the course of the chemical reaction the ligand could easily approach the heme group. These structures do not indicate any preference for 2- or 4-hydroxylation, as in several poses the distances between the iron ion in the heme group and carbon atoms at 2 and 4 positions (C_2 and C_4) in estrone are similar. This result was also supported by docking estrone into various other CYP3A4 X-Ray structures. In all cases a large variety of docked poses was obtained (data not shown). These results are also in line with our observation that re-docking of the co-crystallized inhibitor to CYP3A4 yielded numerous poses

In the case of the CYP1 family members much smaller number of poses have been predicted by AutoDock Vina, which originates from the reduced active site size (see Figures S2-S4 in the SI). The structures can be classified into two major groups: either ring A or ring D lies close to the heme group, but in both sets of structures a well-defined hydrophobic interaction is observed between the sterane skeleton and a conserved phenylalanine residue (CYP1B1 – Phe231, CYP1A1 – Phe224, CYP1A2 – Phe226) of the α-helix restricting the active site. Out of these structures only those are physiologically relevant in which ring A is closer to the heme group, as it is known that CYP1 members either 2- or 4-hydroxylate estrogen derivatives, selected structures are shown in Fig. 4. It is worth noting, though, that estrogens can also be 16-hydroxylated on ring D, especially by CYP3A4/5 isoforms.[72]

CYP1A1 and CYP1A2 bind estrone in a quite similar way, due to their very similar structure. In these isoforms C_2 of estrone is clearly closer to the heme iron than C_4. This result is in accordance with the experimental data which states that estrone is primarily 2-hydroxylated by CYP1A1 and CYP1A2. Interestingly, in the case of CYP1B1 the automated docking yielded only binding modes where ring A of estrone is far away from the reaction centrum, while ring D was close to it. This is in sharp contrast to previous experimental data which clearly suggests that CYP1B1 4-hydroxylates estrone, thus ring A should be close to the heme group instead of ring D. Therefore, we looked for alternative binding modes of estrone in CYP1B1 by manually rotating the ligand in the active site. Despite the narrow space we

found a suitable position: in this case C_4 is clearly closer to the heme group than C_2. This binding mode resembles the findings of an earlier homology modelling study.[25]

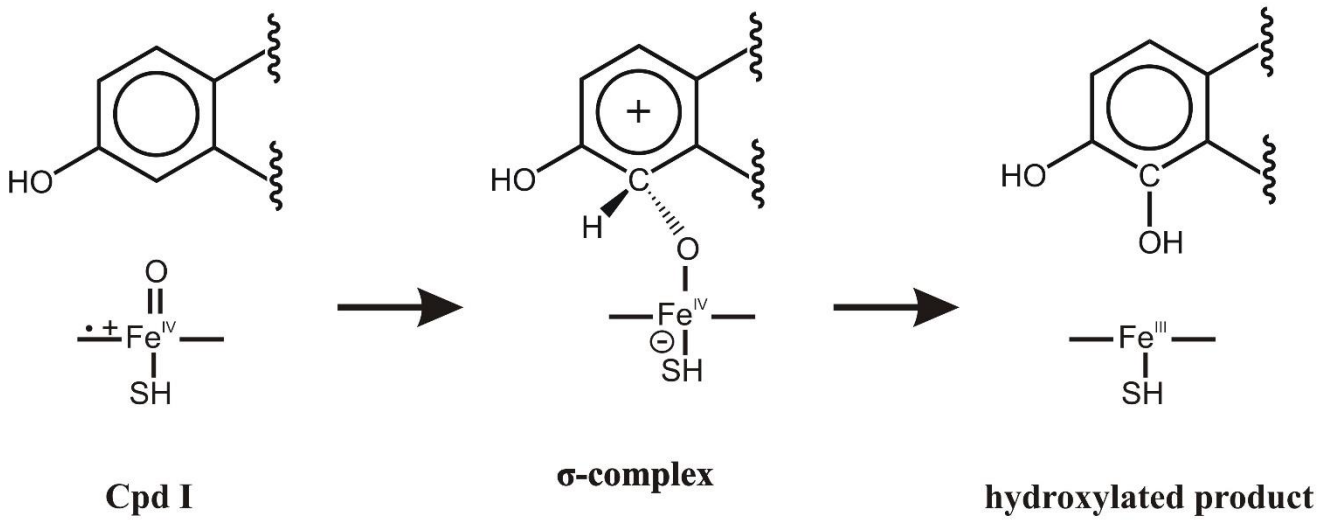
From the results of the structural investigation of the active sites of the four isoforms and of the docking, we conclude that they are in accordance with the experiment and support the hypothesis that CYP1A1 and CYP1A2 bind estrone (and most likely other estrogen derivatives) in a mode that facilitates 2-hydroxylation, in CYP1B1 C_4 of estrone is closer to the heme group than C_2, thus 4-hydroxylation seems to be preferred, while in CYP3A4 due to the large active site cavity both 2- and 4-hydroxylations may occur.

As both the intrinsic reactivity of the various sites in a molecule and the preferred binding mode of the substrate can influence the CYP-mediated metabolism of compounds,[65] in the next part of our paper we investigated the intrinsic reactivity of C_2 and C_4 carbon atoms in three estrogen derivatives (see Chart 1).

3.2. Calculations using Compound I of CYP enzymes.

The mechanism of aromatic carbon oxidation has been studied in detail by de Visser and Shaik.[73] The ultimate oxidant species of CYP enzymes is Compound I, which shows multistate reactivity: it can carry out oxidation reactions both in the quartet and doublet spin states.[74] In the first step of the reaction a tetrahedral σ-complex is formed which can show either cationic or radical character and from these complexes various routes exist to yield the final product (see Scheme 2), which could be either a hydroxylated product or an epoxide or more rarely a keto compound. In all cases the rate-determining step of the reaction is the formation of the σ-complex and the pathway involving the lowest energy transition state is the cationic pathway in the doublet state.[73] For this reason we studied only this pathway (see the collected NBO charges and spin densities in Table S3 and Table S4 in the Supporting Information). First we determined the transition states for addition of Compound I to C_2 and C_4 of the model compounds, and also the σ-complexes that are formed. From the relative energies of the transition states of the 2- and 4-hydroxylated pathways one can draw conclusion on the relative product ratios using

Arrhenius theory; the route with lower activation energy contributes more significantly to product formation and yields more products.



Scheme 2. Schematic representation of the proposed reaction mechanism of aromatic hydroxylation by CYP enzymes. The reaction proceeds via a tetrahedrally arranged σ-complex with a cationic character, which easily rearranges to yield the final product.

In Table 2 we have collected the relative energies and in Table 3 selected geometrical parameters of the transition states and σ–complexes for all compounds. Depending on the relative orientation of the heme ring and the estrogen ring two possibilities exist: the estrogen may approach the heme group in a sideon or in a faceon orientation (see Figure 5). In general, the faceon transition state is slightly later than the sideon transition state: this can be seen from the shorter substrate C-compound I oxygen distance, from the more elongated iron-oxygen distance, and from the slightly larger charge accumulated on the ligand in the transition state (see Table S3 in the Supporting Information). If we compare the structure of the transition states for 2- and 4-hydroxylations we can observe for all compounds and both faceon and sideon orientations that the Fe-O-C angle is larger for 4-hydroxylation by 4-5 degrees than for 2-hydroxylation, but most of the geometrical parameters are very similar for 2- and 4-hydroxylations.

Table 2. Activation energies (**$\Delta E^{\#}$**) and reaction energies (**ΔE**) for the studied addition reactions (leading to the σ-complex) in kcal/mol at the B3LYP/6-311+G* / B3LYP/6-311+G*+ZPE / B3LYP/6-311+G*+D3 dispersion correction level of theory.

Compound		OMe		OPhe		Compound I			
						sideon		faceon	
		ΔE	$\Delta E^{\#}$	ΔE	$\Delta E^{\#}$	ΔE	$\Delta E^{\#}$	ΔE	$\Delta E^{\#}$
A	2 hyd	-3.0/-0.8/-2.9	7.2/8.1/5.8	22.6/23.3/19.9	25.4/25.1/21.6	-9.5/-9.3/-15.3	18.7/18.4/12.6	-7.2/-6.7/-15.7	19.9/19.7/10.8
	4 hyd	-3.9/-1.6/-4.5	6.4/7.2/4.3	20.4/21.1/16.6	25.3/25.0/20.5	-10.9/-10.4/-19.3	19.2/19.1/10.9	-5.6/-5.1/-15.6	20.1/20.1/9.4
B	2 hyd	-2.7/-0.6/-2.6	7.5/8.3/6.0	23.1/23.7/20.4	25.9/25.6/22.1	-8.3/-8.3/-12.5	19.3/19.1/15.1	-6.2/-5.8/-13.0	20.2/20.0/12.5
	4 hyd	-2.7/-0.6/-2.6	7.5/8.3/6.0	20.5/21.3/16.8	25.3/25.1/20.5	-10.1/-9.8/-16.8	19.4/19.3/12.8	-5.1/-4.7/-13.0	20.2/20.0/11.1
C	2 hyd	-3.6/-1.6/-3.1	8.7/9.3/7.5	21.1/21.8/18.6	26.8/26.5/23.0	-4.0/-4.0/-8.5	21.2/20.8/17.4	-1.5/-1.1/-7.9	22.1/21.8/15.5
	4 hyd	-9.2/-6.8/-9.0	4.9/5.6/3.7	17.4/18.4/14.0	21.9/21.8/17.6	*	17.3/17.3/11.8	-9.5/-8.8/-16.8	18.6/18.5/11.4

* we could not locate the σ-complex with the right structure.

Table 3. Selected geometrical data for the transition states and σ−complexes (in parentheses) of the studied addition reactions at the B3LYP/BSI level. Bond distances are given in Å and angles in degrees.

Com-pound		**OMe**			**OPhe**			**Compound I**					
								sideon			**faceon**		
		C_s-O[#]	**O-C_r**[#]	**C_sOC_r**[#]	**C_s-O**[#]	**O-C_r**[#]	**CsOC_r**[#]	**C_s-O**[#]	**O-Fe**[#]	**C_sOFe**[#]	**C_s-O**[#]	**O-Fe**[#]	**CsOFe**[#]
A	**2 hyd**	1.940	1.399	112.91	1.786	1.351	117.09 (117.62)	2.072	1.698	125.85	1.926	1.728	138.53
		(1.458)	(1.425)	(114.48)	(1.490)	(1.388)		(1.360)	(2.023)	(122.64)	(1.355)	(2.054)	(127.88)
	4 hyd	1.957	1.399	112.74	1.785	1.353	117.97	2.040	1.705	130.54	1.934	1.730	141.09
		(1.462)	(1.426)	(114.38)	(1.473)	(1.380)	(121.42)	(1.353)	(2.059)	(127.13)	(1.346)	(2.113)	(132.31)
B	**2 hyd**	1.938	1.399	112.96	1.784	1.352	117.06	2.059	1.701	126.01	1.917	1.729	138.22
		(1.458)	(1.426)	(114.50)	(1.489)	(1.388)	(117.60)	(1.359)	(2.025)	(122.45)	(1.354)	(2.056)	(127.60)
	4 hyd	1.959	1.400	112.80	1.785	1.353	117.69	2.047	1.704	129.80	1.930	1.730	140.63
		(1.461)	(1.427)	(114.35)	(1.470)	(1.381)	(121.40)	1.353	2.059	127.09	(1.345)	(2.120)	(132.41)
C	**2 hyd**	1.963	1.399	113.17	1.804	1.349	117.70	2.046	1.706	127.84	1.921	1.727	138.51
		(1.451)	(1.427)	(114.49)	(1.459)	(1.384)	(121.16)	(1.351)	(2.018)	(121.40)	(1.348)	(2.051)	(125.58)
	4 hyd	2.013	1.396	112.49	1.843	1.345	117.50	2.099	1.693	129.26	1.957	1.729	140.51
		(1.451)	(1.428)	(114.83)	(1.482)	(1.391)	(118.54)	(*)	(*)	(*)	(1.354)	(2.111)	(134.12)

[#] C_s: carbon in the substrate, C_r: carbon in the methoxy or phenoxy groups see labels in Figure 5.

[*] we could not locate the σ-complex with the right structure.

In general the barriers are calculated to be slightly, by about 1 kcal/mol lower when the two fragments approach each other in a sideon orientation, but if dispersion is taken into account the trend is reversed. In the faceon orientation, the heme and steroid rings interact more intimately resulting in more favourable dispersion interactions leading to a lower barrier by about 2-3 kcal/mol than in the sideon orientation. In all cases dispersion seems to be very significant, in accordance with previous results,[75] it is about 4-6 kcal/mol in the sideon orientation and about 8-10 kcal/mol in the faceon orientation. In all cases it is more pronounced for the 4-hydroxylation pathway than for 2-hydroxylation.

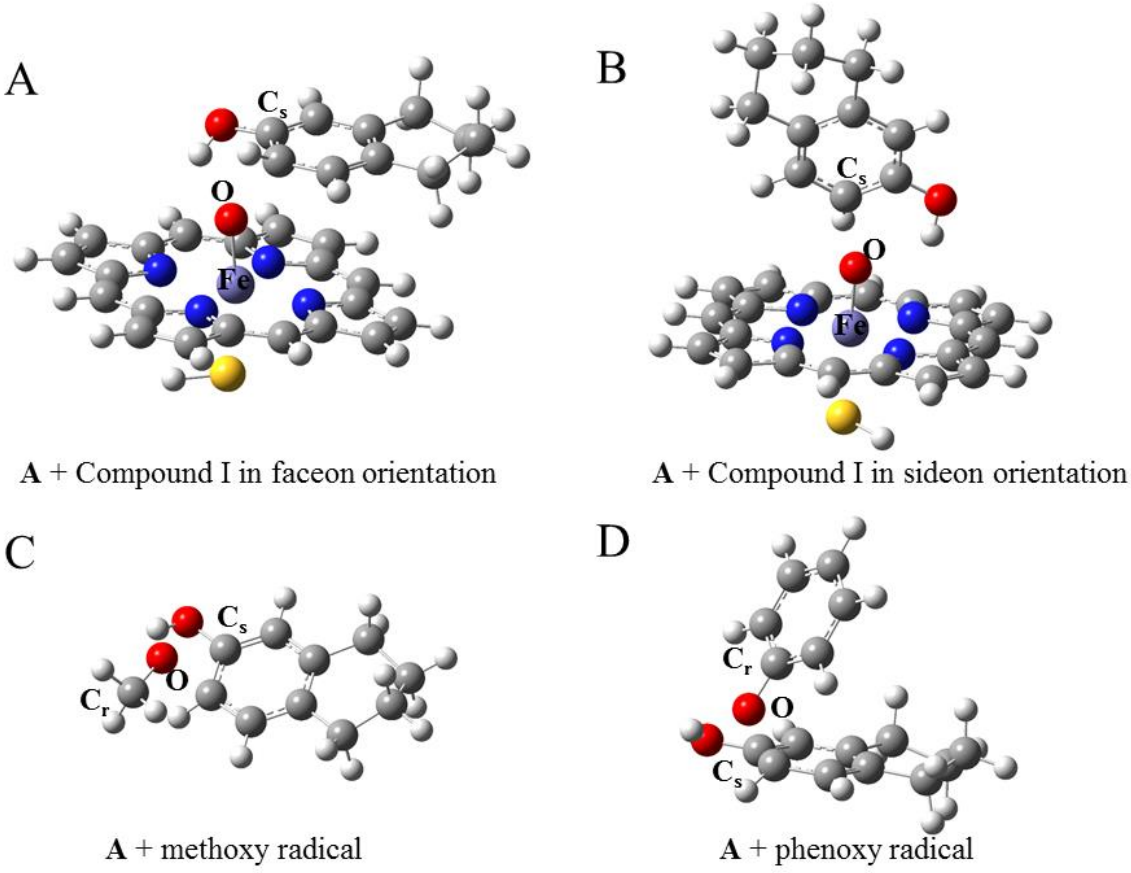


Figure 5. Illustration of the various transition states located in the present work through the example of 2-hydroxylation of **A**, model of estrone. Atom labels used in Table 3 are also shown (C_s: substrate carbon, C_r: radical carbon). Color coding: red: oxygen, white: hydrogen, dark blue: iron, light blue: nitrogen, grey: carbon: yellow: sulfur.

In the case of **A**, model of estrone 2-hydroxylation has a slightly lower barrier than 4-hydroxylation (19.7 vs 20.1 kcal/mol, and 18.4 and 19.1 kcal/mol in the faceon and sideon orientations, respectively without dispersion) in full accordance with the experimental results which showed a 9:1 ratio of 2- and 4-hydroxylated products. If dispersion is taken into account the barrier is slightly lower for 4-hydroxylation (10.8 vs 9.4 kcal/mol and 12.6 and 10.9 kcal/mol in the faceon and sideon orientations, respectively for 2- vs 4 -hydroxylation). Nevertheless, the two barriers in both cases are in accordance with the experimentally observed mixture of products. In the case of **B**, model of equilin the barriers for the faceon orientation with and without dispersion are for 2-hydroxylation: 20.0, 12.5 kcal/mol and for 4-hydroxylation 20.0 and 11.1 kcal/mol. The barrier difference between 2- and 4-hydroxylation is smaller than in the case of estrone (see Table 4), in complete accordance with the experimental fact that a similar ratio of 2- and 4-hydroxylated products was observed. In the case of **C**, model of equilenin the barriers for the faceon orientation with and without dispersion are 2-hydroxylation: 21.8, 15.5 and 4-hydroxylation 18.5 and 11.4 kcal/mol). These data indicate on the one hand the slightly increased barrier of 2-hydroxylation compared to estrone and equilin and the slightly decreased barrier of 4-hydroxylation, and these two effects together account for a significantly increased thermodynamic preference for 4-hydroxylation in the case of equilenin.

Using Arrhenius theory at room temperature and supposing a similar reaction mechanism (which is most likely the case) 4 kcal/mol difference between the activation energies of two routes should lead to a product ratio of about 1:1000. This result may explain why in the case of 2-napthol and equilenin only 4-hydroxylated products have been observed.[29,30]

It is worth looking at the results of quantum chemical and docking calculations at the same time as both the orientation of the ligand in the active site and the tendency of the various poses to be oxidized will determine the final product ratio. One important conclusion based on the QM calculations is that both faceon and sideon orientation of the ligands could lead to successful reactions, as the transition states in the two orientations have similar energies. Looking at the obtained poses in CYP1 family members estrone is oriented in a more sideon-like manner, while in CYP3A4 both sideon- and faceon-like transition

states could occur due to the large cavity. Earlier we showed in the case of dextromethorphan metabolism by CYP2D6 that it is a prerequisite of metabolic reaction to occur that a transition state with favorable orientation could be formed, and it is not enough if a site is close-by to the heme group and has a high tendency to be oxidized.[54] Based on the docking and QM results we suggest that in CYP1 family members the most important factor that determines the regioselectivity of the oxidation reaction is the orientation of the ligand in the active sites and the possibility for forming a favorable, low-energy transition state, while in the case of CYP3A4 the most important factor will be the intrinsic reactivity of the different sites, as there is a lot of space for the reorientation of the ligand in the active site which could allow the most reactive site to be oxidized. Therefore, we suggest that all of estrone, equilin and equilenin could be 2-hydroxylated by CYP1A1 and 1A2, and 4-hydroxylated by CYP1B1. In contrast, in the case of CYP3A4 we expect a change from mixed 2- and 4-hydroxylation of estrone to exclusive 4-hydroxylation of equilenin. Our result imply that in the case of the baboon and hamster liver microsome preparation that was used for determining the ratios of 2- and 4-hydroxylated products an isoform similar to CYP3A4 with large cavities should have been responsible for the metabolism of the majority of the substrate, and that the intrinsic reactivities of C_2 and C_4 were the most important factors influencing the final product ratios.

3.3. Calculations using methoxy and phenoxy radical as model of CYP enzymes.

In the pharmaceutical industry a lot of effort is invested into both experimental and computational methods to predict the most likely metabolites of drug candidates at an early stage.[76,77] One of the methodologies that has been suggested to reduce the computational cost is to use a methoxy radical (or phenoxy)[26] to mimic the effect of Compound I. Naturally, these methods only give us information on the intrinsic reactivity of the two sites of the enzymes, which could be relevant in the case of large active sites (such as that of CYP3A4), in other cases, as we noted above, the active site architecture might be the most important factor influencing the regioselectivity of the oxidation reaction. Although the phenoxy and methoxy models seem to be most adaptive for the investigation of reactions which involve hydrogen abstraction, we decided to check their performance. Therefore, the lowest energy transition states were

determined between the substrates and methoxy/phenoxy radicals, as well as the formed complexes. Relevant structural data is shown in Table 3, relative energies can be found in Table 2, and natural populations and spin densities in the Supporting Information.

The obtained activation energies differ a lot from those obtained with Compound I: with the methoxy radical the barriers are lower by about 10 kcal/mol, and with the phenoxy radical they are higher by about 5-6 kcal/mol. However, in order to predict the relative ratios of 2- and 4-hydroxylation pathways, not the barrier heights themselves, but the differences between the two barriers are the most relevant (see Table 4). It can be observed that all of our CYP models (Compound I, methoxy and phenoxy radical) correctly reproduce the experimentally observed trend: as the conjugation between ring A and B increases the barrier of 2-hydroxylation becomes higher than that of 4-hydroxylation (in the order estrone<equilin<equilenin). Furthermore, the activation energy differences are reasonably close to the experimental value, and are in the same order for both the methoxy and the phenoxy radical as for Compound I. This suggests that when the reactivities of similar sites are compared, these simpler models might be a good solution for the investigation of the reactivity of molecules.

Table 4. Calculated activation energy differences (in kcal/mol) between the 2- and 4-hydroxylation pathways using various model systems. Fukui indices calculated at C_2 and C_4 of the estrogen derivatives at the B3LYP/6-311+G*+ZPE and B3LYP/6-311+G*+ZPE+D3 correction (in parenthesis) level of theory.

	exp. Ratio[27]	ΔE#2-ΔE#4				f-		f+		f0	
Compound		calc from exp.	Compound I faceon	OMe	OPhe	C2	C4	C2	C4	C2	C4
A	9.000[£]	-1.28*	-0.4 (1.4)	0.9 (0.5)	0.1 (1.1)	0.096	0.045	0.161	0.037	0.128	0.041
B	2.300[£]	-0.48*	0 (1.4)	1 (1.6)	0.5 (1.6)	0.082	0.040	0.221	0.018	0.152	0.029
C	<0.300[£,#]	>0.70*	3.3 (4.1)	3.7 (3.8)	4.7 (5.4)	0.006	0.154	0.111	0.121	0.058	0.138

[£] measured for the estrogen derivatives, estrone, equilin and equilenin, not for the model compounds to which the calculated data refer to.

[#] only the 4-hydroxylated product was observed

[*] this value was calculated based on the experimental product ratios at 293 K. It is only a very crude estimation, due to the fact that several CYP isoform may have been present in the reaction mixture which may have different binding affinities for the ligands, may have different reaction rates for 2. vs 4-hydroxylations of the various substrates.

3.4. Reactivity indices

Reactivity indices have been developed to offer a simple method to predict the reactivity compounds or of the various sites in a compound. We calculated the condensed Fukui functions for electrophilic, nucleophilic and radical attack for each carbon atom of compounds **A**, **B** and **C** based on eq. (1)-(3) (see Table 4). Compound I of CYP enzymes is a radical cation, which is a very strong electrophile and oxidizing agent. The mechanism of substrate oxidation may follow cationic or radical pathways.[78] For this reason one would expect that either the Fukui function for nucleophilic attack (i.e. in which the reacting agent is electrophilic, thus the nucleophilicity of C_2 and C_4 are expected to determine their relative reactivity) or for radical attack could provide information on the most likely sites to be attacked by Compound I. In the case of **A**, where 2-hydroxylation is more prevalent according to experiment than 4-hydroxylation all three types of Fukui functions predict larger reactivity of C_2 than of C_4. For **B**, the Fukui function for nucleophilic attack (f^-) show closer values to each other than in the case of **A** in accordance with the similar reactivity of C_2 and C_4 observed in the experiment. In contrast both f^+ and f^0 predict C_2

to be much more reactive against a nucleophile.[29,30] In the case of **C** both f^- and f^0 predict C_4 to be more reactive towards an electrophile or a radical, while against a nucleohile similar reactivities of the two sites are expected. This is in line with experimental data, 2-napthol and equilenin are only 4-hydroxylated.[29,30]

Although these reactivity indices are just simple methods to check for the intrinsic reactivity of the molecules, but it is interesting to see that indeed the nucleophilic and radical Fukui functions correctly predict the trends observed for 2 vs 4-hydroxylations by the strong electrophilic oxidizing species of CYP enzymes. It may be well-worth examining in the future whether these trends hold true for further estrogen derivatives.

4. Conclusions

In this paper we used a combination of modelling tools (docking and quantum chemical calculations) to gain insight into the factors that govern the 2- and 4-hydroxylation pathways of three estrogens, estrone, equilin and equlenin. We found that in most cases docking provides reasonable insight into the docking mode of estrone into the active site of human estrogen-metabolizing CYP enzymes (1A1, 1A2, 1B1, 3A4). The results suggested that in CYP1A1 and CYP1A2 C_2 of estrone is closer to the heme group than C_4, while in the case of CYP1B1 the opposite was found, which is in accordance with the experimentally observed reactivities. Due to the large active site of CYP3A4 the possibility exists for metabolism at both sites. Quantum chemical calculations on the 2- and 4- hydroxylation routes of models of estrone, equilin, and equilenin clearly indicated that C_4 becomes more nucleophilic with increasing conjugation between rings A and B, thus it will be much more prone to be oxidized by Compound I of CYP enzymes. This trend is also suggested by the radical and nucleophilic condensed Fukui functions, and proved by determining the activation energies of 2- and 4-hydroxylation pathways. Our results also strongly suggest that in the experiments using hepatic microsomal preparations for the study of estrone, equilin and equilenin metabolism an isoform similar to human CYP3A4 significantly contributed to product formation and influenced the final product ratios.

Funding Sources: AL thanks the financial support of the Richter Gedeon Talentum Foundation. JO was supported by EU Marie Curie ERG Fellowship (Project "Oestrometab"), the Bolyai János Research Scholarship and by NKFIH Grant No. 115503. BK was supported by OTKA Grant No. 108721. Financial Support from New Széchenyi Plan (Project ID: TAMOP-4.2.2/B-10/1-2010-0009) is also acknowledged.

Supporting Information: Binding modes and ligand binding affinities of estrone in the active sites of CYP enzymes, ligand binding affinities and RMSD values (compared to the X-Ray structure) of re-docked co-crystallized inhibitors, cartesian coordinates of all located transition states and σ-complexes, atomic charges and spin densities calculated for various fragments. This material is available free of charge via the Internet at http://pubs.acs.org."

Abbreviations:

CYP – cytochrome P450 enzyme

D3 – Grimme's D3 dispersion correction

f_A^+ – Fukui function for a nucleophilic attack

f_A^- – Fukui function for electrophilic attack

f_A^0 – Fukui function for a radical attack on the system

HRT – hormone replacement therapy

NBO – natural bond orbital analysis

NPA – natural population analysis

QM – quantum mechanical/quantum chemical

RMSD – root mean square deviation

ZPE zero-point vibrational energy

REFERENCES

1. Bernhardt, R. Cytochromes P450 as versatile biocatalysts. *J. Biotechnol.* **124,** 128–145 (2006).

2. Montellano, P. R. O. de. Cytochrome P450-activated prodrugs. *Future Med. Chem.* **5,** 213–228 (2013).

3. Chen, W. Y. Exogenous and endogenous hormones and breast cancer. *Best Pract. Res. Clin. Endocrinol. Metab.* **22,** 573–585 (2008).

4. Henderson, B. E., Ross, R. & Bernstein, L. Estrogens as a cause of human cancer: the Richard and Hinda Rosenthal Foundation award lecture. *Cancer Res.* **48,** 246–53 (1988).

5. Thompson, J. A. *et al.* Oxidative metabolism of butylated hydroxytoluene by hepatic and pulmonary microsomes from rats and mice. *Drug Metab. Dispos.* **15,** 833–40 (1987).

6. Liehr, J. G. Genotoxic effects of estrogens. *Mutat. Res. Genet. Toxicol.* **238,** 269–276 (1990).

7. Pike, M. C., Krailo, M. D., Henderson, B. E., Casagrande, J. T. & Hoel, D. G. 'Hormonal' risk factors, 'breast tissue age' and the age-incidence of breast cancer. *Nature* **303,** 767–770 (1983).

8. Beral, V. & Collaborators, M. W. S. Breast cancer and hormone-replacement therapy in the Million Women Study. *Lancet* **362,** 419–427 (2003).

9. Yager, J. D. & Davidson, N. E. Estrogen Carcinogenesis in Breast Cancer. *N Engl J Med* **354,** 270–282 (2009).

10. Cavalieri, E. *et al.* Catechol estrogen quinones as initiators of breast and other human cancers: Implications for biomarkers of susceptibility and cancer prevention. *Biochim. Biophys. Acta - Rev. Cancer* **1766,** 63–78 (2006).

11. Russo, J. & Russo, I. H. The role of estrogen in the initiation of breast cancer. *J. Steroid Biochem. Mol. Biol.* **102,** 89–96 (2006).

12. Li, J. J., Li, S. A., Oberley, T. D. & Parsons, J. A. Carcinogenic Activities of Various Steroidal and

Nonsteroidal Estrogens in the Hamster Kidney: Relation to Hormonal Activity and Cell Proliferation. *CANCER Res.* **55,** 4347–4351 (1995).

13. Henderson, B. E. & Feigelson, H. S. Hormonal carcinogenesis. *Carcinogenesis* **21,** 427–433 (2000).

14. Flötotto, T. *et al.* Hormones and Hormone Antagonists: Mechanisms of Action in Carcinogenesis of Endometrial and Breast Cancer. *Horm. Metab. Res.* **33,** 451–457 (2001).

15. Revankar, C. M., Cimino, D. F., Sklar, L. A., Arterburn, J. B. & Prossnitz, E. R. A transmembrane intracellular estrogen receptor mediates rapid cell signaling. *Science* **307,** 1625–30 (2005).

16. X-D Song, R., Fan, P., Yue, W., Chen, Y. & Santen, R. J. Role of receptor complexes in the extranuclear actions of estrogen receptor a in breast cancer. *Endocr. Relat. Cancer* **13,** 3–13 (2006).

17. Rogan, E. G. *et al.* Relative imbalances in estrogen metabolism and conjugation in breast tissue of women with carcinoma: potential biomarkers of susceptibility to cancer. *Carcinogenesis* **24,** 697–702 (2003).

18. Castagnetta, L. A. M. *et al.* Tissue content of hydroxyestrogens in relation to survival of breast cancer patients. *Clin. Cancer Res.* **8,** 3146–55 (2002).

19. Bolton, J. L. & Thatcher, G. R. J. Potential Mechanisms of Estrogen Quinone Carcinogenesis. *Chem. Res. Toxicol.* **21,** 93–101 (2008).

20. Tompkins, L. M. & Wallace, A. D. Mechanisms of cytochrome P450 induction. *J. Biochem. Mol. Toxicol.* **21,** 176–81 (2007).

21. Zhu, B. T. & Lee, A. J. NADPH-dependent metabolism of 17β-estradiol and estrone to polar and nonpolar metabolites by human tissues and cytochrome P450 isoforms. *Steroids* **70,** 225–244 (2005).

22. Tsuchiya, Y., Nakajima, M. & Yokoi, T. Cytochrome P450-mediated metabolism of estrogens and its regulation in human. *Cancer Lett.* **227,** 115–124 (2005).

23. Parl, F. F., Dawling, S., Roodi, N. & Crooke, P. S. Estrogen Metabolism and Breast Cancer: a Risk Model. *Ann. N. Y. Acad. Sci.* **1155,** 68–75 (2009).

24. Liehr, J. G. & Ricci, M. J. 4-Hydroxylation of estrogens as marker of human mammary tumors. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 3294–6 (1996).

25. Itoh, T., Takemura, H., Shimoi, K. & Yamamoto, K. A 3D Model of CYP1B1 Explains the Dominant 4-Hydroxylation of Estradiol. *J. Chem. Inf. Model.* **50,** 1173–1178 (2010).

26. Olsen, L., Rydberg, P., Rod, T. H. & Ryde, U. Prediction of Activation Energies for Hydrogen Abstraction by Cytochrome P450. *J. Med. Chem.* **49,** 6489–6499 (2006).

27. Rydberg, P., Ryde, U. & Olsen, L. Prediction of Activation Energies for Aromatic Oxidation by Cytochrome P450. *J. Phys. Chem. A* **112,** 13058–13065 (2008).

28. Dai, Z.-R. *et al.* A Mechanism-Based Model for the Prediction of the Metabolic Sites of Steroids Mediated by Cytochrome P450 3A4. *Int. J. Mol. Sci.* **16,** 14677–14694 (2015).

29. Purdy, R. H., Moore, P. H., Williams, M. C., Goldzieher, J. W. & Paul, S. M. Relative rates of 2- and 4-hydroxyestrogen synthesis are dependent on both substrate and tissue. *FEBS Lett.* **138,** 40–44 (1982).

30. Sarabia, S. F., Zhu, B. T., Kurosawa, T., Tohma, M. & Liehr, J. G. Mechanism of Cytochrome P450-Catalyzed Aromatic Hydroxylation of Estrogens. *Chem. Res. Toxicol.* **10,** 767–771 (1997).

31. Kapucuoglu, N. *et al.* Expression of CYP3A4 in human breast tumour and non-tumour tissues. *Cancer Lett.* **202,** 17–23 (2003).

32. Shaik, S., Kumar, D., de Visser, S. P., Altun, A. & Thiel, W. Theoretical Perspective on the

Structure and Mechanism of Cytochrome P450 Enzymes. *Chem. Rev.* **105,** 2279–2328 (2005).

33. Shaik, S. *et al.* P450 Enzymes : Their Structure , Reactivity , and Selectivity s Modeled by QM / MM Calculations. *Chem. Rev.* **110,** 949–1017 (2010).

34. Krámos, B., Menyhárd, D. K. & Oláh, J. Direct Hydride Shift Mechanism and Stereoselectivity of P450nor Confirmed by QM/MM Calculations. *J. Phys. Chem. B* **116,** 872–85 (2012).

35. Krámos, B. & Oláh, J. The mechanism of human aromatase (CYP 19A1) revisited: DFT and QM/MM calculations support a compound I-mediated pathway for the aromatization process. *Struct. Chem.* **26,** 279–300 (2015).

36. Krámos, B. & Oláh, J. Enolization as an Alternative Proton Delivery Pathway in Human Aromatase (P450 19A1). *J. Phys. Chem. B* **118,** 390–405 (2014).

37. Lonsdale, R., Oláh, J., Mulholland, A. J. & Harvey, J. N. Does compound I vary significantly between isoforms of cytochrome P450? *J. Am. Chem. Soc.* **133,** 15464–74 (2011).

38. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7,** 539–539 (2014).

39. Li, W. *et al.* The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res.* **43,** W580–W584 (2015).

40. McWilliam, H. *et al.* Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res.* **41,** W597–W600 (2013).

41. Wang, A., Savas, U., Stout, C. D. & Johnson, E. F. Structural Characterization of the Complex between α-Naphthoflavone and Human Cytochrome P450 1B1. *J. Biol. Chem.* **286,** 5736–5743 (2011).

42. Walsh, A. A., Szklarz, G. D. & Scott, E. E. Human Cytochrome P450 1A1 Structure and Utility in

Understanding Drug and Xenobiotic Metabolism. *J. Biol. Chem.* **288,** 12932–12943 (2013).

43. Sansen, S. *et al.* Adaptations for the Oxidation of Polycyclic Aromatic Hydrocarbons Exhibited by the Structure of Human P450 1A2. *J. Biol. Chem.* **282,** 14348–14355 (2007).

44. Sevrioukova, I. F. & Poulos, T. L. Pyridine-Substituted Desoxyritonavir Is a More Potent Inhibitor of Cytochrome P450 3A4 than Ritonavir. *J. Med. Chem.* **56,** 3733–3741 (2013).

45. Webb, B. & Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinforma.* **47,** 5.6.1-32 (2014).

46. Fiser, A., Do, R. K. & Sali, A. Modeling of loops in protein structures. *Protein Sci.* **9,** 1753–73 (2000).

47. Li, H., Robertson, A. D. & Jensen, J. H. Very fast empirical prediction and rationalization of protein pKa values. *Proteins* **61,** 704–21 (2005).

48. Bas, D. C., Rogers, D. M. & Jensen, J. H. Very fast prediction and rationalization of pKa values for protein-ligand complexes. *Proteins Struct. Funct. Bioinforma.* **73,** 765–783 (2008).

49. Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3 : Consistent Treatment of Internal and Surface Residues in Empirical p K a Predictions. *J. Chem. Theory Comput.* **7,** 525–537 (2011).

50. Søndergaard, C. R., Olsson, M. H. M., Rostkowski, M. & Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J. Chem. Theory Comput.* **7,** 2284–95 (2011).

51. Wishart, D. S. *et al.* DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **34,** D668-72 (2006).

52. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new

scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31,** 455–61 (2010).

53. Brändén, G., Sjögren, T., Schnecke, V. & Xue, Y. Structure-based ligand design to overcome CYP inhibition in drug discovery projects. *Drug Discov. Today* **19,** 905–911 (2014).

54. Sevrioukova, I. F. & Poulos, T. L. Dissecting Cytochrome P450 3A4–Ligand Interactions Using Ritonavir Analogues. *Biochemistry* **52,** 4474–4481 (2013).

55. Kaur, P., Chamberlin, A. R., Poulos, T. L. & Sevrioukova, I. F. Structure-Based Inhibitor Design for Evaluation of a CYP3A4 Pharmacophore Model. *J. Med. Chem.* **59,** 4210–4220 (2016).

56. Sevrioukova, I. F. & Poulos, T. L. Structural and Mechanistic Insights into the Interaction of Cytochrome P4503A4 with Bromoergocryptine, a Type I Ligand. *J. Biol. Chem.* **287,** 3510–3517 (2012).

57. Sevrioukova, I. F. & Poulos, T. L. Interaction of human cytochrome P4503A4 with ritonavir analogs. *Arch. Biochem. Biophys.* **520,** 108–116 (2012).

58. Sevrioukova, I. F. & Poulos, T. L. Structure and mechanism of the complex between cytochrome P4503A4 and ritonavir. *Proc. Natl. Acad. Sci.* **107,** 18422–18427 (2010).

59. Ekroos, M. & Sjogren, T. Structural basis for ligand promiscuity in cytochrome P450 3A4. *Proc. Natl. Acad. Sci.* **103,** 13682–13687 (2006).

60. Bathelt, C. M., Zurek, J., Mulholland, A. J. & Harvey, J. N. Electronic Structure of Compound I in Human Isoforms of Cytochrome P450 from QM/MM Modeling. *J. Am. Chem. Soc.* **127,** 12900–12908 (2005).

61. Frisch, M. J. *et al.* Gaussian 09, Revision A.01. (2009).

62. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.*

132, 154104 (2010).

63. Grimme, S., Ehrlich, S. & Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **32,** 1456–1465 (2011).

64. Bathelt, C. M. *et al.* Mechanism and structure–reactivity relationships for aromatic hydroxylation by cytochrome P450. *Org. Biomol. Chem.* **2,** 2998–3005 (2004).

65. Olah, J., Mulholland, A. J. & Harvey, J. N. Understanding the determinants of selectivity in drug metabolism through modeling of dextromethorphan oxidation by cytochrome P450. *Proc. Natl. Acad. Sci.* **108,** 6050–6055 (2011).

66. Geerlings, P., Proft, F. De & Langenaeker, W. Conceptual Density Functional Theory. *Chem. Rev.* **103,** 1793–1874 (2003).

67. Reed, A. E., Weinstock, R. B. & Weinhold, F. Natural population analysis. *J. Chem. Phys.* **83,** 735 (1985).

68. Mendez, F. & Gazquez, J. L. Chemical Reactivity of Enolate Ions: The Local Hard and Soft Acids and Bases Principle Viewpoint. *J. Am. Chem. Soc.* **116,** 9298–9301 (1994).

69. Mount, D. *Bioinformatics: Sequence and Genome Analysis*. (Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 2004).

70. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. **42,** 320–324 (2014).

71. Yu, J., Zhou, Y., Tanaka, I. & Yao, M. Roll: a new algorithm for the detection of protein pockets and cavities with a rolling probe sphere. *Bioinformatics* **26,** 46–52 (2010).

72. Huang, Z., Guengerich, F. P. & Kaminsky, L. S. 16Alpha-hydroxylation of estrone by human cytochrome P4503A4/5. *Carcinogenesis* **19,** 867–72 (1998).

73. de Visser, S. P. & Shaik, S. A Proton-Shuttle Mechanism Mediated by the Porphyrin in Benzene Hydroxylation by Cytochrome P450 Enzymes. *J. Am. Chem. Soc.* **125,** 7413–7424 (2003).

74. de Visser, S. P., Shaik, S., Sharma, P. K., Kumar, D. & Thiel, W. Active Species of Horseradish Peroxidase (HRP) and Cytochrome P450: Two Electronic Chameleons. *J. Am. Chem. Soc.* **125,** 15779–15788 (2003).

75. Lonsdale, R., Harvey, J. N. & Mulholland, A. J. Inclusion of Dispersion Effects Significantly Improves Accuracy of Calculated Reaction Barriers for Cytochrome P450 Catalyzed Reactions. *J. Chem. Lett.* **1,** 3232–3237 (2010).

76. Kirchmair, J. *et al.* Predicting drug metabolism: experiment and/or computation? *Nat. Rev. Drug Discov.* **14,** 387–404 (2015).

77. Olsen, L., Oostenbrink, C. & Jørgensen, F. S. Prediction of cytochrome P450 mediated metabolism. *Adv. Drug Deliv. Rev.* **86,** 61–71 (2015).

78. Meunier, B., de Visser, S. P. & Shaik, S. Mechanism of Oxidation Reactions Catalyzed by Cytochrome P450 Enzymes. *Chem. Rev.* **104,** 3947–3980 (2004).