



**QM/MM modelling of the hydroxylation of the androstenedione substrate catalyzed by Cytochrome P450 Aromatase (CYP19A1)**

Journal:	<i>Journal of Computational Chemistry</i>
Manuscript ID:	JCC-14-0643.R2
Wiley - Manuscript type:	Full Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Martí, Sergio; Universitat Jaume I, Departament de Química Física i Analítica Viciano, Ignacio; Universitat Jaume I, Departament de Química Física i Analítica Cstillo, Raquel; Universitat Jaume I, Departament de Química Física i Analítica
Key Words:	Cytochrome P450, Compound I, Aromatase, Hydrogen Abstraction, Androstenedione, CYP19A1, Quantum Mechanics / Molecular Mechanics

SCHOLARONE™  
Manuscripts

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

# QM/MM modelling of the hydroxylation of the androstenedione substrate catalyzed by Cytochrome P450 Aromatase (CYP19A1)

*Ignacio Viciano, Raquel Castillo and Sergio Marti\**

Departament de Química Física i Analítica, Universitat Jaume I, 12071 Castelló, Spain.

\* to whom correspondence should be addressed

[smart@uji.es](mailto:smart@uji.es)

For Peer Review

**Abstract**

CYP19A1 aromatase is a member of the Cytochrome P450 family of heme proteins, and is the enzyme responsible for the final step of the androgens conversion into the corresponding estrogens, via a three-step oxidative process. For this reason, the inhibition of this enzyme plays an important role in the treatment of hormone-dependent breast cancer. The first catalytic subcycle, corresponding to the hydroxylation of androstenedione, has been proposed to occur through a first hydrogen abstraction and a subsequent oxygen rebound step. In present work, we have studied the mechanism of the first catalytic subcycle by means of hybrid QM/MM methods. The inclusion of the protein flexibility has been achieved by means of Free Energy Perturbation techniques, giving rise to a free energy of activation for the hydrogen abstraction step of 13.5 kcal/mol. The subsequent oxygen rebound step, characterized by a small free energy barrier (1.5 kcal/mol), leads to the hydroxylated products through a highly exergonic reaction. In addition, an analysis of the primary deuterium kinetic isotopic effects, calculated for the hydrogen abstraction step, reveals values ( $\sim 10$ ) overpassing the semiclassical limit for the C-H, indicating the presence of a substantial tunnel effect. Finally, a decomposition analysis of the interaction energy for the substrate and cofactor in the active site is also discussed. According to our results, the role of the enzymatic environment consists of a transition state stabilization by means of dispersive and polarization effects.

**Keywords:** Cytochrome P450, Compound I, Aromatase, Hydrogen Abstraction, Androstenedione, CYP19A1, Quantum Mechanics / Molecular Mechanics (QM/MM).

## 1. Introduction

Cytochrome P450 (CYP or P450) is a superfamily of heme-thiolate-containing metalloenzymes, composed by a large number of families and subfamilies, which belong to the three domains of life. Currently there are more than 11500 CYP named, 57 of which are members of the *homo sapiens* species.<sup>[1]</sup> These enzymes play an important role in different biological processes in some organisms such as plants, mammalian, bacteria, fungi, archaea or protists.<sup>[2-3]</sup> P450s can metabolize a wide variety of both endogenous and exogenous substrates, being responsible of the metabolism for a large amount of pharmaceuticals<sup>[4]</sup> (~75%). The most common reactions catalyzed by CYP enzymes involve oxygen insertions, including hydroxylation of unactivated C-H bonds, alkene epoxidations, N- and O-dealkylations, heteroatom oxidations such as sulfoxidations, etc.<sup>[2, 5]</sup>

Human Aromatase (CYP19A1) is a member of the superfamily of cytochrome P450 enzymes and is located in a number of tissues that include gonads, adrenal glands, ovaries, placenta, testes, adipose tissue and numerous sites in the brain.<sup>[6-7]</sup> This enzyme is responsible for a key step in the biosynthesis of steroid hormones, and more specifically it is involved in the final step of the conversion of androgens (androstenedione (ASD), testosterone and 16- $\alpha$ -hydroxytestosterone) into the corresponding estrogens (estrone, 17- $\beta$ -estradiol and estriol, respectively).<sup>[8]</sup> This conversion consists of three consecutive oxidation steps of the 19-methyl group of the androgens, which is removed as formic acid during the course of the reaction, resulting in aromatization of ring A of these steroidal substrates (see scheme 1).<sup>[9]</sup> First and second steps are believed to proceed through two sequential hydroxylations of the 19-methyl group of ASD to produce 19-*gem*-diol species that then undergoes dehydration to yield the aldehyde intermediate (19-oxo-ASD). The third step, which is still under discussion, corresponds to the C<sub>10</sub>-C<sub>19</sub> bond cleavage process (lyase reaction).<sup>[10]</sup> The entire process of aromatization consumes a total of three moles of NADPH and three moles of oxygen.<sup>[2]</sup>

Aromatase is the only known enzyme responsible for the biosynthesis of estrogens from androgens in vertebrates. The estrogens are the primary female sex hormones, being important for sexual and reproductive development in humans, mainly in woman. However, these steroidal hormones also stimulate the growth of hormone-receptor-positive breast cancer cells playing an important role in the development of hormone-dependent breast cancer. The inhibition of this enzyme, and hence cessation of estrogen production, provides one of the first molecular targets for rational drug development in the treatment of breast cancer and of other hormone-responsive cancers.<sup>[11-12]</sup> For this reason, the development of potent and selective aromatase inhibitors (AIs) is one of the greatest challenges in the field of endocrinology regarding the breast cancer therapy, as well as a goal to achieve in the pharmaceutical industry.

The knowledge of the structural nature of the enzyme and its mechanism of action provide useful information for designing new AIs and optimal drugs to treat breast cancer. One of the best ways to explore the structure of an enzyme as well as its reaction mechanisms is through the use of computational techniques

1 such as molecular dynamics (MD) simulations and hybrid quantum mechanics/molecular mechanics  
2 (QM/MM) methods. Although the cytochrome P450 superfamily of enzymes has been extensively studied  
3 using computational methods in some of its most common isoforms,<sup>[13]</sup> CYP450 aromatase has not been  
4 under theoretical study for a long time, partly because of the lack of a good crystallographic model. This fact  
5 prompted a number of homology models based on other P450 isoforms and site-directed mutagenesis data  
6 were proposed.<sup>[14]</sup> The crystal structure of aromatase purified from human placenta in complex with its  
7 natural substrate ASD (PDB code 3EQM) was determined by Ghosh and coworkers<sup>[15]</sup> in 2009. This crystal  
8 structure shows how the natural substrate is accommodated in the active site of this enzyme and provides a  
9 good initial molecular structure on which to apply computational tools. Since the publication of this  
10 structure, some mechanistic studies on aromatase catalysis using QM/MM methods have been published in  
11 recent years.<sup>[16] [17]</sup> The first oxidation step of the 19-methyl group of ASD has been studied experimentally  
12 many times,<sup>[18-26]</sup> however, no QM/MM study has been carried out for this catalytic subcycle so far, thus  
13 remaining the mechanism still unsolved to this day.

21 This first oxidation step, in which ASD is hydroxylated to 19-hydroxy-ASD, is thought to proceed through  
22 the hydrogen abstraction/oxygen rebound mechanism, originally proposed by Groves et al., as occurs in  
23 other P450s.<sup>[27]</sup> This mechanism, consists of an initial hydrogen atom abstraction of the substrate by means  
24 of the reactive species, a high valent ferryl-oxo heme complex (Compound I, Cpd I), leading to the formation  
25 of the alkyl radical intermediate and the iron-hydroxo complex. Subsequently, an alkyl (or OH) reorientation  
26 is produced to facilitate the final oxygen rebound mechanism, where a recombination of the alkyl radical  
27 with the iron-bound hydroxyl radical leads to the formation of the corresponding hydroxylated product (see  
28 scheme 2). The nature of the reactive species located in the active site of CYP enzymes, has been extensively  
29 discussed in the literature by both theoretical<sup>[13]</sup> and experimental<sup>[28]</sup> studies. Currently, the most accepted  
30 oxidant is the Cpd I, an iron IV oxo-porphyrin cation radical ( $\text{Por}^+-\text{Fe}^{\text{IV}}=\text{O}$ ), whose ability to activate inert  
31 C-H bonds has been amply demonstrated.<sup>[29-33]</sup> Even so, due to the elusiveness of this compound and its high  
32 reactivity, making it very difficult to detect and capture, other oxidants such as the ferric hydroperoxo  
33 complexes ( $\text{Por}-\text{Fe}^{\text{III}}-\text{OOH}^+$ , Cpd 0) and perferryl based species ( $\text{Por}-\text{Fe}^{\text{V}}=\text{O}$ ) have been proposed.<sup>[28, 34-39]</sup>  
34 However, Rittle and Green<sup>[30]</sup> succeeded preparing Cpd I on CYP119 P450 cytochrome in a high yield, being  
35 able to spectroscopically and kinetically characterize it, settling the matter on the hydroxylating agent in  
36 P450 enzymes. In addition, in a recent study,<sup>[40]</sup> Khatri et al. concluded that the Cpd I is the reactive  
37 intermediate not only in the first two steps of aromatization by aromatase (hydroxylations) but also in the  
38 third (and last) aromatization step.

51 The Compound I has a radicalary nature and possesses three singly occupied orbitals. Two of them are the  
52  $\pi_{xz}^*$  and  $\pi_{yz}^*$  orbitals of the Fe-O moiety, while the third one is delocalized between the  $a_{2u}$  orbital of the  
53 porphyrin and the  $\pi_s$  of the sulfur atom. A ferromagnetic coupling of the three electrons provides a high-spin  
54 quartet state ( $S=3/2$ ), whereas an antiferromagnetic one leads to a low-spin doublet state ( $S=1/2$ ). As a result  
55 of this dual spin nature of the Cpd I, Shaik and coworkers<sup>[41]</sup> proposed a two-state reactivity mechanism  
56  
57  
58  
59  
60

(TSR) for the hydroxylation and the epoxidation of alkanes, in which product distribution of the reaction is determined by the combination of both spin states.

We herein present a theoretical study for the first oxidation step of androgen substrate ASD via cytochrome CYP19A1 aromatase. In this oxidation step, which involves the hydroxylation of the 19-methyl group of ASD to 19-hydroxy-ASD, the radical rebound step as well as the Cpd I as the oxidizing species have been proposed. The two potential energy surfaces (PESs), corresponding to the two possible spin configurations, doublet and quartet, have been explored by means of QM(DFT)/MM methods. In addition, the activation free energies of the biochemical process have been evaluated introducing Free Energy Perturbation (FEP) techniques, thus taking into account the conformational space of the protein. Also, semiclassical kinetic isotope effects (KIE) of the hydrogen abstraction step from ASD 19-methyl group have been reported. Finally, an exhaustive analysis of the binding nature of both substrate and cofactor has been carried out by means of the decomposition of the interaction potential energy along the hydrogen abstraction step.

## 2. Computational Methods

### 2.1. System Setup and MD Simulations

The initial geometry used in our calculations was obtained from the X-ray crystal structure of the human placental aromatase cytochrome P450, in complex with its natural substrate ASD, solved at 2.9 Å of resolution (PDB code 3EQM).<sup>[15]</sup> The original pentacoordinated heme B cofactor found in the PDB file was modeled into an hexacoordinated Iron-Oxo porphyrin Cpd I species, using the atom positions suggested in the literature.<sup>[15]</sup> All the hydrogen atoms were added using the NAMD 2.8 program,<sup>[42]</sup> according to the empirical pKa predictions rendered by PROPKA web interface.<sup>[43]</sup> The standard protonation states in solution were found except for Asp-309, which exhibited a large pKa displacement (7.7). This value indicated a weak acidic character and therefore Asp-309 was protonated at pH 7. Histidine residues were protonated as follows: His-171 was doubly protonated, His-(62, 105, 111, 325, 475 and 480) were singly protonated at  $\epsilon$  position, His-(109, 128, 402 and 459) were singly protonated at  $\delta$  position. A total of 4 counter ions (Cl<sup>-</sup>) were placed into optimal electrostatic positions around the enzyme in order to electroneutralize the excess of positive charge (+4), fulfilling the electroneutrality of the system. The model was placed in a prerelaxed orthorhombic box of water molecules with dimensions of 90x80x80 Å<sup>3</sup>, erasing all those water molecules with an oxygen atom lying less than 2.8 Å from any heavy atom. The resulting model consisted of 452 residues of amino acids, the modeled Cpd I, the substrate ASD, 35 crystallographic water molecules, 4 counterions and 16542 water molecules of the solvation box.

The solvated model was minimized and then equilibrated by means of classical MD at 300 K, using the NVT ensemble and the Langevin–Verlet integrator. The MD was run in NAMD for 20 nanoseconds (ns) with a step size of 1 femtosecond (fs) using the OPLS-AA<sup>[44]</sup> force field and the TIP3P<sup>[45]</sup> water model. The nonbonding interactions were treated in all calculations by periodic boundary conditions with the Particle

Mesh Ewald convention, via a force-switch function with a cutoff distance in the range 14-16 Å. No harmonic restraints were applied between atoms and/or fragments during the simulations.

## 2.2. QM and QM/MM Calculations

Starting structure used in QM/MM calculations was extracted from one of the last snapshots obtained in MD simulations based on observation criteria. In order to study the chemical reactivity, the QM region of the model was defined to include the substrate, the Cpd I and the axial Cys-437 ligand (comprising the S<sub>γ</sub>, the C<sub>β</sub> and two H<sub>β</sub> atoms), giving rise to a total of 124 atoms (see Figure 1a). The link atom formalism was applied to satisfy the valence of the QM region, due to presence of the classical bond partitioning in the Cys-437 amino acid (C<sub>α</sub>-C<sub>β</sub> bond). Because of the complexity of the enzymatic model, a truncation scheme was implemented during all hybrid QM/MM calculations. In this scheme, the lists of interaction for the QM atoms (incorporating all atoms within 20 Å from any atom of both the substrate and heme groups) are defined at the very beginning and remain constant during the calculations. In this way, all residues further than 20 Å from either the Cpd I or the substrate were kept frozen, during the simulations (8638 atoms from a total of 57252).

Since the study and characterization of a transition state (TS) usually make use of the Hessian matrix, its evaluation is not feasible for enzymatic systems due to the large amount of degrees of freedom. One way to overcome this problem is to divide the full coordinates space into two different subsets: the control space, which incorporates all those atoms or molecules involved in the reaction process (usually the QM atoms); and the complementary space comprising the rest of the system. Therefore, the optimization of the model system can be expressed as the combination of iterations in both subsets: at each step of the control space Hessian guided optimization (based on the Baker<sup>[46-47]</sup> algorithm), the rest of the system is kept fully relaxed merely using gradient vectors (using the L-BFGS<sup>[48]</sup> procedure). This strategy is also known as the micro/macroiteration method, and allows taking advantage of efficient optimization algorithms.<sup>[49]</sup> Furthermore, a dual QM:Charge/MM scheme<sup>[49-50]</sup> was adopted, where the total QM/MM energy of the system is expressed using different terms, depending on which coordinate space is being optimized. This choice is justified by the fact that a typical complementary space optimization (corresponding to the macroiterations) typically needs a large number of gradient evaluations at each control space movement (or microiteration), and thus some approximations must be introduced in order to speedup calculations. In this case, the QM atoms are reduced to frozen classical charges during the macroiterations, thereby using different expressions for the energy depending on the current iteration, as shown in following equations:

$$E_{micro} = E_{MM} + \langle \Psi | \hat{H}_o | \Psi \rangle + \sum \langle \Psi | \frac{q_{MM}}{r} | \Psi \rangle = E_{MM} + E_{QM,o} + E_{QM,int} \quad (1)$$

$$E_{macro} = E_{MM} + \sum \frac{q_{QM}^{fit} q_{MM}}{r_{QM,MM}} \quad (2)$$

1 The term “ $E_{MM}$ ”, common in both equations, is the energy of the MM region, including the Lennard-Jones  
2 interaction between QM and MM atoms. The terms “ $E_{QM,o}$ ” and “ $E_{QM,int}$ ” included in equation (1),  
3 correspond to the quantum energy of the QM atoms using a high-level method (such as *ab initio* or density  
4 functional based methods) under a classical environment described by point charges ( $q_{MM}$ ). The term  
5 “ $E_{QM,o}$ ” refers to the selected gas-phase Hamiltonian operator acting on the polarized wave function,  
6 meanwhile the term “ $E_{QM,int}$ ” involves the electrostatic interaction among this wave function and the  
7 classical point charges. On the other hand, in equation (2) the gas-phase quantum energy calculation has  
8 been removed, and the QM/MM electrostatic interaction is replaced by a Coulombic expression, making use  
9 of electrostatic potential fitted charges (Merz-Kollman<sup>[51-52]</sup>) for the QM atoms.

10 In this way, the TSs were localized by means of this micro/macroiterations scheme, and were characterized  
11 by hessian inspection. Afterwards, minimum energy paths (MEP) were then traced down to their  
12 corresponding minima, to ensure that the structures really connect reactants and intermediates. Once the  
13 reaction paths were obtained, the Gibbs free energy was evaluated at the stationary points, under the rigid-  
14 rotor and the harmonic-oscillator (RRHO) approximations, taking into account both zero point energy (ZPE)  
15 and thermal corrections.

16 Moreover, a small model of the system (see Figure 1b) was built in order to study the hydrogen abstraction  
17 step, performing gas-phase calculations (QM-only model). This model, consisting of a reduced version of  
18 both Cpd I and ASD, was used to contrast some results obtained for the enzymatic model.

19 The unrestricted Kohn-Sham formalism with the B3LYP<sup>[53-54]</sup> hybrid density functional was used to describe  
20 the QM region of the enzyme aromatase as well as the quantum atoms of the QM-only model. It has been  
21 shown that the DFT and DFT/MM methodologies provide a reasonably good description of the doublet and  
22 quartet ground states of the Cpd I.<sup>[55-56]</sup> In particular, the B3LYP functional has become the preferred choice  
23 for carrying out the modeling of the CYP enzymes since is well known to give reasonable relative spin state  
24 energies as well as to perform well for a number of other properties.<sup>[13, 57]</sup> The LACVP\* basis set (B1),<sup>[58]</sup>  
25 consisting of the combination of the 6-31G(d) basis for all the atoms except for the iron, which is  
26 represented by the Lanl2dz effective core potential (ECP), was used for the exploration of the PES.  
27 Additionally, in order to obtain more accurate energies and frequency numbers, a larger basis set (B2) was  
28 adopted to perform single point calculations on the stationary points previously allocated with the B1 basis  
29 set. This B2 basis set corresponds to the combination of the Lanl2tz+ ECP for the iron and the 6-311G(d,p)  
30 basis for the rest of the atoms. The rest of the enzyme and the water molecules were described by the OPLS-  
31 AA and TIP3P force fields, respectively.

32 All the electronic calculations were performed using the Gaussian03<sup>[59]</sup> family of programs, whereas the  
33 fDYNAMO<sup>[60]</sup> library was used to describe the molecular mechanics during the hybrid QM/MM  
34 simulations.

### 35 2.3 Free Energy Perturbation Calculations



Free Energy Perturbation techniques were introduced in order to evaluate the free energies of activation of the chemical process. The use of this method, based on statistical mechanics, allows taking into account the conformational space of the enzyme, obtaining accurate free-energy surfaces and solvent effects. For this purpose, different energetically equispaced geometries (windows) were selected from the previously traced reaction paths, being 54 for the doublet spin state and 75 for the quartet state. In each window, the QM region (including the link atom) was kept frozen during the calculations. Following the micro/macro scheme presented before, after each QM calculation performed (see equation 1), 100 steps of classical MD were carried out with a time step of 1 fs (see equation 2), under the same conditions as during the equilibration. This procedure was repeated 200 times resulting in a total of 20 picoseconds (ps) per window. Finally, the coordinates of the QM region were swapped between consecutive windows (for the last 100 structures) in order to estimate the free energy changes, using the following expression:

$$\Delta F_{i \rightarrow j} = -RT \ln \left\langle e^{-\left(\frac{H_j - H_i}{RT}\right)} \right\rangle_i \quad (3)$$

In equation 3,  $H_i$  corresponds to the QM/MM potential energy obtained during the  $i$ th window of MD. The term  $H_j$  refers to the QM/MM potential energy, once the coordinates of the atoms for the control space have been exchanged ( $j=i+1$  for forward or  $j=i-1$  for backwards), under the same geometry of the classical environment ( $i$ ). Finally, the triangular brackets denote an average performed over the  $i$ th window. Once the free energy barrier is calculated using Equation 3, the ZPE correction term is included as the difference between the ZPEs of both final and initial states (reactant and transition states for the activation energy) by means of QM/MM frequency calculations.

#### 2.4 Kinetic Isotope Effects (KIE) Calculations

Semiclassical kinetic isotope effects were calculated for the hydrogen abstraction step in both QM-only model and the enzymatic system (as an average of 10 window obtained from FEP calculations for both the TS and reactants).

$$KIE_E = \frac{k_H}{k_D} = \exp \left[ \frac{\Delta G_D^\ddagger - \Delta G_H^\ddagger}{RT} \right] \quad (4)$$

Subsequent corrections to the Eyring semiclassical KIE were incorporated by means of the Wigner correction<sup>[61]</sup> in order to estimate the Tunnel effects on these KIEs:

$$KIE_W = KIE_E \cdot \left( \frac{Q_{W,H}}{Q_{W,D}} \right) \quad ; \quad Q_W = 1 + \frac{1}{24} \left( \frac{h\nu^\ddagger}{k_B T} \right)^2 \quad (5)$$

where  $\nu^\ddagger$  is the imaginary frequency at the saddle point,  $k_b$  is Boltzmann's constant and  $h$  is the Planck's constant. This choice seems to be appropriate to the tunneling corrections applied to rate constants for which the values of transmission coefficients are small to moderate ( $\leq 2$ ).<sup>[62-63]</sup> The CAMVIB/CAMISO<sup>[64-65]</sup>

1 programs, based on the RRHO approximations, without scaling the vibrational frequencies, were used to  
2 calculate KIEs.

3  
4 Since the analysis of the bond orders (BO) may provide relevant information about the geometrical  
5 similarity in the TS structures, which is directly related with the KIE values, the C<sub>19</sub>-H<sub>19</sub> and H<sub>19</sub>-O formal  
6 BO were calculated. For this purpose, Natural Bond Orbital (NBO) calculations were performed using the  
7 Gaussian 03 package at the UB3LYP/B1 level in order to estimate the NAO Wiberg bond indices.<sup>[66]</sup>  
8  
9

### 10 11 12 13 **3. Results and Discussion**

14  
15  
16 The results presented in this paper correspond to the hydroxylation of the ASD 19-methyl group to 19-  
17 hydroxy-ASD (see scheme 2) catalyzed by the cytochrome P450 aromatase. Both the initial hydrogen  
18 abstraction step, and the subsequent oxygen rebound step were studied in the active site of the enzyme, in  
19 the doublet and the quartet spin states, using the B3LYP[B1]:ESP/MM scheme presented before. Table 1  
20 shows the Gibbs free energies for the QM-only model (using the B1 basis set), and for the enzymatic system  
21 (using both B1 and B2 basis set). In addition, the free energies of activation obtained from FEP in the  
22 enzymatic model for both basis sets are also presented.  
23  
24  
25

26 The QM-only model gas-phase calculations, exhibit almost degenerated reactants at both spin states. The  
27 Gibbs activation energies, as well as the corresponding imaginary frequencies, for the hydrogen abstraction  
28 step show also similar values, leading to the formation of the intermediates through an endergonic process.  
29 The comparison of these values with those obtained in the enzymatic environment (using the same B1 basis  
30 set), displays a modest breakdown of the reactants degeneracy by 1.4 kcal/mol. The subsequent Gibbs  
31 activation free energies for this step are lowered by 3.4 kcal/mol in the doublet and 1.3 kcal/mol in the  
32 quartet, when comparing to the gas-phase model. This modest catalytic effect obtained from the PES  
33 analysis suggests a mild participation of the enzymatic environment in the chemical process. The  
34 intermediates that follow the hydrogen abstraction step give place to the oxygen rebound mechanism by  
35 means of a small barrier (~3 kcal/mol), leading to the formation of the hydroxylated products through a  
36 highly exergonic process (-32 and -43 kcal/mol for the doublet and quartet states respectively).  
37  
38

39 The inclusion of the conformational space effects (enzyme flexibility), by means of the FEP methodology,  
40 dramatically reduces the free energy of activation for the hydrogen abstraction step. The obtained value of  
41 13.5 kcal/mol for both spin states appears to be as a more realistic prediction for this biochemical  
42 reaction.<sup>[30, 67]</sup> At the same time, the free energy difference between reactants in both spin states obtained  
43 with the FEP is increased when compared with the PES results, despite that both spin states are accessible  
44 during the dynamics. Once the intermediate species have been formed, the oxygen rebound step takes place  
45 after a previous reorientation process (see scheme 2).  
46  
47

48 For the doublet electronic state, the rotation has a small free energy barrier of 2.1 kcal/mol leading to a  
49 rotation intermediate (-3.3 kcal/mol) from which the rebound step takes place with a barrier of 2.3 kcal/mol.  
50 This behavior is different in the quartet state, where the rotation has a free energy barrier of 2.2 kcal/mol  
51 leading to a rotation intermediate roughly thermoneutral (0.7 kcal/mol) from which the rebound takes place  
52  
53  
54  
55  
56  
57  
58  
59  
60

with a barrier of 2.7 kcal/mol. Finally, the inclusion of the ZPE energy term reduces the overall barrier for the rebound step to 1.5 and 2.8 kcal/mol, for the doublet and quartet states respectively. As can be noted, the free energies of the product species are not reported in Table 1, since we have focused on obtaining only the free energies of activation for the biochemical process with the FEP methodology.

The effect of the basis set has been also inspected by means of the incorporation of a larger basis set (B2), thus performing single point calculations of the structures obtained in the former PES (B1 basis set). The new Gibbs free energies obtained depict a decrease in the activation energy of the hydrogen abstraction around 2-3 kcal/mol, while the remaining stationary points show similar values of relative energy. Additionally, the basis set correction has been included in the FEP calculation, but only in those terms that can be factored out of the average, in order to avoid the large amount of expensive QM(B2) calculations:

$$\Delta F \approx -RT \ln \left\langle e^{-\frac{\Delta E_{int}^{B1}}{RT}} \right\rangle + \Delta E_{vac}^{B2} + \Delta ZPE^{B2} \quad (6)$$

In equation 6, the interaction energy term ( $\Delta E_{int}^{B1}$ ) has been previously calculated using the B1 basis set, while both the gas-phase energy ( $\Delta E_{vac}^{B2}$ ) and the ZPE ( $\Delta ZPE^{B2}$ ) terms correspond to the expensive calculations with the large basis set. The obtained results show a free energy of activation for the hydrogen abstraction lower than the corresponding to the B1 basis set (11.4 and 12.0 kcal/mol for the doublet and quartet states respectively). This trend, associated to the use of the B2 basis set, is not observed in the rebound step, where the activation barriers remain virtually unchanged (1.3 and 3.3 kcal/mol for the doublet and quartet state respectively).

Primary deuterium ( $^2\text{H}$ ) KIE values of the  $\text{H}_{19}$  abstraction step for the two models (QM-only and the enzymatic system) were determined using the frequency data obtained with the B1 basis set. This basis set was used in order to compare the results for the QM-only model and the enzymatic system (obtained as an average of the FEP structures), due to the complexity of the latter.

Inspection of the computed KIEs presented in Table 2 show values around 7 in both models and in both spin states (doublet and quartet). The fact that both models provide almost identical values (very close to semiclassical Eyring limit,  $k_{\text{H}}/k_{\text{D}} \approx 7$ ) suggests that this KIE value is intrinsic to the reaction and is not affected by the enzymatic environment. Regarding the slight difference in values between different spin states, we can invoke the effect of the symmetry of the TS since it is well known that the KIE value depends on the geometry adopted in this stationary point. Thus, the analysis of the distances among atoms involved in the H-transfer coordinate ( $d_{(\text{C}_{19}-\text{H}_{19})}$  and  $d_{(\text{H}_{19}-\text{O})}$ ) for the TS might provide a picture of the relationship between the symmetry and the KIE values. The geometrical parameters obtained for the TS species along this coordinate in the enzymatic system are:  $d_{(\text{C}_{19}-\text{H}_{19})}=1.399\text{\AA}$ ,  $d_{(\text{H}_{19}-\text{O})}=1.133\text{\AA}$  and  $\text{C}_{19}-\text{H}_{19}-\text{O}=171.2^\circ$  for the doublet, and  $d_{(\text{C}_{19}-\text{H}_{19})}=1.394\text{\AA}$ ,  $d_{(\text{H}_{19}-\text{O})}=1.134\text{\AA}$  and  $\text{C}_{19}-\text{H}_{19}-\text{O}=169.6^\circ$  for the quartet. These distances and angles obtained along the H-transfer coordinate are practically the same in both spin states, and hence denote a great geometrical similarity between the TS species obtained in both states. In fact, this behavior of geometrical similarity is also borne out by the results obtained from NBO analysis performed on TS structures. The values provided by Wiberg bond order (BO) indices on the atoms involved in the H-transfer

coordinate are:  $BO_{(C_{19}-H_{19})}=0.343$  and  $BO_{(H_{19}-O)}=0.471$  for the doublet and  $BO_{(C_{19}-H_{19})}=0.344$  and  $BO_{(H_{19}-O)}=0.470$  for the quartet state.

In this way, due to the strong parallel drawn between TS structures, we can find similar imaginary frequencies for the hydrogen abstraction step in doublet (1606*i*) and quartet (1645*i*) spin states, and even when the deuterium atom replaces the transferred hydrogen atom (1209*i* and 1236*i* for doublet and quartet respectively).

This analogous behavior between doublet and quartet spin states along the H-transfer coordinate, can also be found in the case of the QM-only model. In this model, the geometrical parameters obtained for the TS species along this coordinate are:  $d_{(C_{19}-H_{19})}=1.409\text{\AA}$ ,  $d_{(H_{19}-O)}=1.141\text{\AA}$  and  $C_{19}-H_{19}-O=166.4^\circ$  for the doublet, and  $d_{(C_{19}-H_{19})}=1.413\text{\AA}$ ,  $d_{(H_{19}-O)}=1.143\text{\AA}$  and  $C_{19}-H_{19}-O=164.1^\circ$  for the quartet. The Wiberg bond order indices obtained are:  $BO_{(C_{19}-H_{19})}=0.350$  and  $BO_{(H_{19}-O)}=0.468$  for the doublet,  $BO_{(C_{19}-H_{19})}=0.348$  and  $BO_{(H_{19}-O)}=0.469$  for the quartet. Finally, the imaginary frequencies for the hydrogen abstraction step are: 1601*i* in doublet and 1713*i* in quartet spin states, and when the deuterium atom replaces the transferred hydrogen atom are 1200*i* and 1277*i* for doublet and quartet respectively. It is noteworthy that both the geometrical parameters and the bond order indices as well as imaginary frequency numbers obtained are very similar between the QM-only model and the enzymatic system, and therefore, it is not surprising that the KIE results are similar in both models.

Corrections for barrier tunneling were determined by Wigner approximation and provided values in the range 1.4 – 1.5 in both models and spin states. Since the Wigner correction yields values significantly greater than unity, it suggests the involvement of a significant amount of tunneling in this hydrogen abstraction step. Indeed, the corrected KIE values increase up to values around 10 in both systems and spin states, thus exceeding the semiclassical Eyring limit (see Table 2). It is worth pointing out that tunneling effect has been reported in previous theoretical<sup>[68-73]</sup> and experimental<sup>[30, 38, 74-77]</sup> studies for this kind of hydrogen atom abstractions.

The Mulliken atomic spin densities obtained for the hydroxylation step using the B2 basis set are reported in Table 3. These results reveal the presence of three unpaired electrons in the reactant species: two electrons with parallel spin are located on the FeO moiety of the Cpd I forming a local triplet ( $\rho_{FeO} \sim 2.10$ ) and a third electron is located between the  $\pi_s$  orbital at the sulfur atom of the Cys-437 and the  $a_{2u}$  orbital of the porphyrin ring. The sign of the latter (beta or alpha) determines either the doublet ( $\rho_{Por+SCH_3} = -1.13$ ) or quartet ( $\rho_{Por+SCH_3} = 0.95$ ) nature of the system respectively. The results in Table 3 show that the porphyrin  $\pi$  cation is predominantly of  $A_{2u}$  character (~80%) in both doublet and quartet spin states, with a significant contribution of  $[\Pi]_s$  state, leading to a “green” Cpd I instead of the “red” one. As long as the first reaction step takes place, the hydrogen  $H_{19}$  is transferred to the oxygen of the Cpd I from the  $C_{19}$  leading to the formation of an alkyl radical on this carbon and a hydroxyl group linked to the iron atom. These results can be observed based on the values of the spin density population obtained for intermediate species,  $\rho_{C_{19}} \sim 1.00$  and  $\rho_{O-H_{19}} \leq 0.15$ . Therefore, the unpaired electron of the hydrogen  $H_{19}$  coming from the homolytic C-H cleavage

1 is transferred to the oxygen atom, thus increasing the negative atomic charge on this atom (see Table S1 of  
2 supplementary material).

3  
4 It is worth noting that the iron-hydroxo complex formed in intermediate species was found in two different  
5 electromeric configurations for the quartet spin state ( ${}^4\text{Fe}^{\text{III}}\text{Por}^{*+}$  and  ${}^4\text{Fe}^{\text{IV}}\text{Por}$ ) and only one for the doublet  
6 state ( ${}^2\text{Fe}^{\text{III}}\text{Por}^{*+}$ ). Both electromers found in the quartet state share the same geometry, but differ in the  
7 distribution of electrons on their atoms.<sup>[13, 78]</sup> The  ${}^{2,4}\text{Fe}^{\text{III}}\text{Por}^{*+}$  electromers contain one unpaired electron on  
8 the iron center ( $\rho_{\text{Fe}} = 0.92$  ( ${}^4\text{Fe}^{\text{III}}\text{Por}^{*+}$ ),  $1.02$  ( ${}^2\text{Fe}^{\text{III}}\text{Por}^{*+}$ )) and the other one distributed on both the  $a_{2u}$  orbital  
9 and the sulfur atom ( $\rho_{\text{Por}+\text{SCH}_3} = 0.97$  ( ${}^4\text{Fe}^{\text{III}}\text{Por}^{*+}$ ),  $-1.09$  ( ${}^2\text{Fe}^{\text{III}}\text{Por}^{*+}$ )). The  ${}^4\text{Fe}^{\text{IV}}\text{Por}$  electromer has the two  
10 unpaired electrons located on the iron center ( $\rho_{\text{Fe}} = 2.06$  and  $\rho_{\text{Por}+\text{SCH}_3} = -0.17$ ). However, all the electromeric  
11 situations retain the radical character on the  $\text{C}_{19}$  atom with a singly occupied orbital ( $\rho_{\text{C}_{19}} = 1.03$  ( ${}^2\text{Fe}^{\text{III}}\text{Por}^{*+}$ ),  
12  $1.04$  ( ${}^4\text{Fe}^{\text{III}}\text{Por}^{*+}$ ) and  $1.05$  ( ${}^4\text{Fe}^{\text{IV}}\text{Por}$ )).

13  
14 The oxygen rebound step is characterized by the formation of a new C-O bond, and involves the elimination  
15 of the radical character in the  $\text{C}_{19}$  carbon atom and the formation of a closed shell porphyrin in the product  
16 species. As a result, two different electronic distributions for  $\text{Fe}^{\text{III}}$  are obtained in products defining the  
17 multiplicity state of the molecular system: the doublet state, which have an uncoupled electron on the iron  
18 atom ( $\rho_{\text{Fe}} = 1.34$  and  $\rho_{\text{Por}+\text{SCH}_3} = -0.34$ ) and the quartet state which have three uncoupled electrons on the iron  
19 (and partially in the sulfur axial ligand) ( $\rho_{\text{Fe}} = 2.70$  and  $\rho_{\text{Por}+\text{SCH}_3} = 0.29$ ).<sup>[13]</sup>

20  
21 Table 4 illustrates some of the most relevant geometrical parameters of the substrate and cofactor in the  
22 active site (for more details see Tables S2 and S3 of supplementary material). The values obtained for both  
23 the Fe-S and Fe-O distances in the reactant species, are consistent with those observed experimentally for  
24 the Cpd I<sup>[79]</sup> (2.48 Å and 1.65 Å respectively, using EXAFS spectroscopy). Concerning the evolution of  
25 these distances as the hydrogen abstraction step proceeds, it can be observed that the Fe-O bond is stretched  
26 due to the formation of the new hydroxyl group linked to the iron atom (from  $\sim 1.63\text{Å}$  to  $\sim 1.80\text{Å}$ ), being this  
27 process accompanied by a shortening of the Fe-S distance (from  $\sim 2.70\text{Å}$  to  $\sim 2.50\text{Å}$ ).

28  
29 Furthermore it can be seen that the donor-acceptor distance ( $\text{C}_{19}\text{-O}$ ) remains practically invariant in the TSs  
30 of both doublet and quartet spin states ( $\sim 2.52\text{Å}$ ). This fact can also be observed in the QM-only model (see  
31 Table S3), where the  $\text{C}_{19}\text{-O}$  bond distance in the TSs ( $\sim 2.53\text{Å}$ ) is similar to that found in the enzymatic  
32 system. However, we can observe that for the reactant species obtained for the enzymatic model, the  $\text{C}_{19}\text{-O}$   
33 distance ( $\sim 2.91\text{Å}$ ) is shorter than the one found in the QM-only model ( $\sim 3.51\text{Å}$ ). This observation suggests  
34 that one of the possible roles of the aromatase active site consists of the pre-organization of the reactant state  
35 towards the TS along the reaction coordinate.

36  
37 Regarding the  $\text{C}_{19}\text{-H}_{19}\text{-O}$  angle, it shows its maximum value in the geometries of the TSs corresponding to  
38 the hydrogen abstraction step. This angle takes values of  $\sim 170$  degrees in the TS of the enzymatic system  
39 and  $\sim 165$  degrees in the TS of the QM-model. The fact that this angle is virtually linear in the TS structures  
40 is what is expected for the hydrogen abstraction step, being this value closer to the linearity in the enzymatic  
41 system than in gas-phase model.

1 Previous to the oxygen rebound step, a reorientation of the hydroxyl group takes place in order to prepare  
2 the formation of the new C<sub>19</sub>-O bond. This rotation can be observed by following the change that occurs in  
3 the H<sub>19</sub>-O-Fe-N<sub>α</sub> dihedral angle between intermediates and TS of the rebound step (around 31° for doublet  
4 and 47° for quartet), together with the shortening of the C<sub>19</sub>-O bond (from ~2.94Å to ~2.60Å). Subsequently,  
5 this distance decreases dramatically from TS to products (from ~2.60Å to ~1.42Å) leading to the formation  
6 of a new C<sub>19</sub>-O bond as well as a penta-coordinated iron complex.  
7  
8  
9  
10

11 Regarding the interactions established among the substrate/cofactor and the residues present in the active  
12 site, the following observations can be made: i) there is a large network of hydrogen bonds at the binding  
13 site of the heme propionate chains involving four charged Arg (115, 145, 375 and 435) and a Trp-141  
14 residues (see Figure 2). These interactions balance the large repulsion of the carboxylate side chains of the  
15 porphyrin ring, facilitating thus an efficient binding site for the heme group in the active site. ii) The  
16 partially charged sulfur atom of the Cys-437 is stabilized by the interaction with three polar peptidic  
17 hydrogens (Ala-438, Gly-439 and Lys-440). In fact, hydrogen bonds have been reported to stabilize the  
18 heme thiolate coordination and regulate the redox potential of the heme iron,<sup>[80]</sup> preventing the elongation of  
19 the Fe-S bond and avoiding the formation of a radical on the sulfur. iii) The binding of the substrate is  
20 accomplished via hydrophobic interactions at the active site with the residues Ile-133, Trp-224, Ala-306,  
21 Thr-310, Val-370, Leu-372, Val-373, Phe-430 and Leu-477. iv) Two hydrogen bonds are established  
22 between the polar moieties of the substrate ASD (carbonyl groups) with the Met-374 and protonated Asp-  
23 309 residues. Previous research suggests that the Asp-309 residue can participate in the enolization reaction  
24 of ASD as a proton donor in the third and last catalytic subcycle of aromatase enzyme<sup>[15]</sup>.  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34

35 The interaction energy between the substrate/cofactor and the enzymatic environment for the hydrogen atom  
36 abstraction step was calculated in the enzymatic system, using the windows corresponding to the reactants  
37 and the TSs obtained from the FEP. This energy has been in turn decomposed into different energetic  
38 contributions as follows: i) the electrostatic term ( $E_{elec}$ ) which accounts for the interaction between the MM  
39 charges of the enzymatic environment and the polarized wave function of the substrate; ii) the polarization  
40 energy ( $E_{pol}$ ) of the wave function due to its changes on-going from gas-phase to the enzymatic active site;  
41 iii) the Lennard-Jones term ( $E_{LJ}$ ) representing the dispersive interactions between the MM and the QM  
42 atoms and iv) the gas-phase energy ( $E_{vac}$ ) of each species. Each of the interaction energy terms were  
43 obtained as an average of both doublet and quartet FEP calculations.  
44  
45  
46  
47  
48

49 As can be seen in Table 5, the electrostatic interaction term ( $E_{elec}$ ) for the doublet TS species shows a weaker  
50 value when compared with the corresponding reactant, leading to an unfavorable contribution to the  
51 activation energy (1.4 kcal/mol). This effect is similar in the quartet state, but less adverse than in the  
52 doublet (1 kcal/mol). On the other hand, it is worth pointing that both polarization ( $E_{pol}$ ) and Lennard-Jones  
53 ( $E_{LJ}$ ) terms stabilize the TSs with respect to the corresponding reactant species, being different the weight of  
54 each term depending on the spin state. In this way, the polarization term in the quartet (-5.9 kcal/mol) is  
55 higher than in the doublet (-2.9 kcal/mol); however, a different behavior occurs for the Lennard-Jones where  
56  
57  
58  
59  
60

1 both the doublet and quartet TSs are stabilized in a similar way (-4.1 and -4.5 kcal/mol respectively).  
2 Furthermore, the  $E_{LJ}$  term can in turn be broken down into the contributions of the substrate ( $\Delta E_{LJ}(ASD)$ ) and  
3 cofactor ( $\Delta E_{LJ}(Cys+heme)$ ), showing that the main contribution to this stabilization comes from the substrate  
4 (ASD) when comparing with the cofactor (Cys+heme), giving rise to values of -2.6 and -4.9 kcal/mol for the  
5 doublet and quartet states respectively.  
6  
7

8  
9 All in all, both the gas-phase barrier and the unfavorable electrostatic TS stabilization are offset by those  
10 contributions from the polarization and Lennard-Jonnes interaction, leading to lower activation free energies  
11 for both spin states. Despite the fact that the free energy of activation is lower for the quartet state, the  
12 reactant is destabilized compared with the doublet, making the overall process to be less affordable.  
13  
14  
15

#### 16 17 18 19 **4. Conclusions**

20  
21 The above study investigates the hydroxylation of the substrate ASD to 19-hydroxy-ASD, occurring during  
22 the first catalytic subcycle of the enzyme aromatase, by means of MD simulations and hybrid QM/MM  
23 calculations. With this study we attempt to improve the understanding of the conversion mechanism of  
24 androgens into estrogens via the enzyme aromatase from the standpoint of theoretical chemistry.  
25  
26

27  
28 In this oxidation step, the radical rebound mechanism (consisting in a first hydrogen abstraction step  
29 followed by an oxygen rebound step) as well as the Cpd I as the oxidizing species have been proposed.  
30

31 According to the results obtained from the analysis performed on the PES, the rate-limiting step corresponds  
32 to the hydrogen abstraction process. The Gibbs free energy barriers for this step (obtained by means of the  
33 RRHO approximation) were found to be slightly lower for the enzymatic system (22.3 kcal/mol) than for the  
34 QM-only model (25.7 kcal/mol), showing a catalytic effect of the enzyme on the biochemical reaction. An  
35 analysis of the  $^2H$ -KIE calculated for this step show almost identical values in both models studied (QM-  
36 only model and enzymatic system), indicating that KIE value is intrinsic to the reaction and is not heavily  
37 affected by the enzymatic environment. Moreover, the results show a slight difference in KIE values  
38 between different spin states, which is explained by the great geometrical similarity found between the TS  
39 species obtained in both states. Once the tunneling correction is included, the KIE values exceed the  
40 semiclassical limit thus denoting a substantial tunnel effect (as has already been reported in some studies  
41 involving C-H abstractions).  
42  
43  
44  
45  
46  
47

48 Furthermore, we have incorporated (for the first time in this P450 isoform) the enzymatic conformational  
49 diversity in the calculations by means of the FEP methodology. Since this methodology takes into account  
50 the enzyme flexibility, it provides a more realistic view of the role played by the enzyme in the biochemical  
51 process. According to the results obtained with this methodology, the free energy barriers for the  
52 hydroxylation step are lower than those calculated by way of the RRHO approximation, being the values  
53 obtained within the range expected for this kind of biological processes. The hydrogen abstraction step takes  
54 place through an endergonic process with an activation barrier of 13.5 kcal/mol, leading to the formation of  
55 an alkyl radical on the substrate and a hydroxyl group linked to the heme cofactor. From these intermediates,  
56  
57  
58  
59  
60

1 a reorientation of the hydroxyl group occurs previous to the oxygen rebound step. The energetics of this  
2 reorientation process depends upon the spin state of the substrate, being the doublet state the more favored.  
3 Thereby, the rebinding of the oxygen befalls via the hydroxyl rotation and the formation of the new C-O  
4 bond. The coupled reorientation-rebound process is characterized by a remarkably low free activation  
5 barrier of 1.5 kcal/mol, giving rise to the formation of the products through a highly exergonic reaction.  
6 Based in these energetic results and the KIEs obtained, we can conclude that the hydroxylation of substrates  
7 via the enzyme aromatase is compatible with the conventional hydrogen abstraction-oxygen rebound  
8 mechanism observed in other P450 isoforms.

9 Finally, the analysis of the decomposition of the free energy barrier (obtained by FEP calculations) in its  
10 different contributions, suggests that the electric field produced by the enzyme during the biochemical  
11 reaction is not directly involved in the stabilization of the transition state when compared with the reactant  
12 species. In fact, it is noteworthy that both the polarization of the wave function and the Lennard-Jones term  
13 are the only ones responsible for such stabilization. These findings are consistent with the kind of chemical  
14 process and/or substrate. Therefore, we can conclude that the main role of the enzyme aromatase during the  
15 hydroxylation of ASD consists of a stabilization of the TS by means of the participation of both dispersive  
16 and polarization effects.

## 27 **5. Acknowledgements**

28 This work was supported by MINECO Project CTQ2012-36253-C03-01, by Generalitat Valenciana  
29 *Prometeo/2009/053* and *PrometeoII/2014/022*, by Universitat Jaume I Project P1.1B2011-23, and by the  
30 HPC-EUROPA2 project (project number: 228398), with the support of the European Community - Research  
31 Infrastructure Action of the FP7. We acknowledge the Servei d'Informàtica of the Universitat Jaume I,  
32 GENCI-CINES and BSC-Marenostrum for providing us with computer capabilities. I.V. thanks Spanish  
33 Ministerio de Ciencia e Innovación for a doctoral grant (CTQ2009-14541-C02). The authors thank V.  
34 Moliner for valuable comments and discussion.

## 43 **Supporting Information**

44 Mulliken Atomic Charges of different atoms or fragments included in the QM part of the enzymatic system  
45 are presented. Also, a summary of the main geometrical parameters obtained in the QM-only model during  
46 the hydrogen abstraction step, as well as a broader presentation of the main distances found in the enzymatic  
47 active site during the hydrogen abstraction/oxygen rebound steps can be found. Finally, the ZPE and the  
48 Gibbs contributions for both spin states, with the B1 basis set, are also reported; as well as the equations  
49 used in the estimation of the FEP uncertainties.



## References

- [1] D. R. Nelson, *Hum. genomics* **2009**, *4*, 59-65.
- [2] P. R. Ortiz de Montellano, *Cytochrome P450 Structure, Mechanism, and Biochemistry*, Third edition ed., Kluwer Academic/Plenum Publishers, New York, **2005**.
- [3] J. S. Lee, R. S. Obach, M. B. Fisher, *Drug metabolizing enzymes : cytochrome P450 and other enzymes in drug discovery and development*, FontisMedia SA and Taylor & Francis Group, USA, **2003**.
- [4] F. P. Guengerich, *Chem. Res. Toxicol.* **2008**, *21*, 70-83.
- [5] F. P. Guengerich, *Chem. Res. Toxicol.* **2001**, *14*, 611-650.
- [6] E. R. Simpson, C. Clyne, G. Rubin, W. C. Boon, K. Robertson, K. Britt, C. Speed, M. Jones, *Annu. Rev. Physiol.* **2002**, *64*, 93-127.
- [7] E. R. Simpson, S. R. Davis, *Endocrinology* **2001**, *142*, 4589-4594.
- [8] Y. Hong, M. Cho, Y. C. Yuan, S. Chen, *Biochem. Pharmacol.* **2008**, *75*, 1161-1169.
- [9] M. Akhtar, J. N. Wright, P. Lee-Robichaud, *J. Steroid Biochem. Mol. Biol.* **2011**, *125*, 2-12.
- [10] M. Akhtar, M. R. Calder, D. L. Corina, J. N. Wright, *Biochem. J.* **1982**, *201*, 569-580.
- [11] R. J. Santen, H. Brodie, E. R. Simpson, P. K. Siiteri, A. Brodie, *Endocr. Rev.* **2009**, *30*, 343-375.
- [12] R. W. Brueggemeier, J. C. Hackett, E. S. Diaz-Cruz, *Endocr. Rev.* **2005**, *26*, 331-345.
- [13] S. Shaik, S. Cohen, Y. Wang, H. Chen, D. Kumar, W. Thiel, *Chem. Rev.* **2010**, *110*, 949-1017.
- [14] D. Ghosh, J. Griswold, M. Erman, W. Pangborn, *J. Steroid Biochem. Mol. Biol.* **2010**, *118*, 197-202.
- [15] D. Ghosh, J. Griswold, M. Erman, W. Pangborn, *Nature* **2009**, *457*, 219-223.
- [16] K. Sen, J. C. Hackett, *Biochemistry* **2012**, *51*, 3039-3049.
- [17] B. Krámos, J. Oláh, *J. Phys. Chem. B* **2013**, *118*, 390-405.
- [18] M. Akhtar, S. J. Skinner, *Biochem. J.* **1968**, *109*, 318-321.
- [19] D. Arigoni, R. Battaglia, M. Akhtar, T. Smith, *J. Chem. Soc., Chem. Commun.* **1975**, *0*, 185-186.
- [20] E. Caspi, T. Arunachalam, P. A. Nelson, *J. Am. Chem. Soc.* **1983**, *105*, 6987-6989.
- [21] E. Caspi, T. Arunachalam, P. A. Nelson, *J. Am. Chem. Soc.* **1986**, *108*, 1847-1852.
- [22] J. Fishman, J. Goto, *J. Biol. Chem.* **1981**, *256*, 4466-4471.
- [23] S. L. Gantt, I. G. Denisov, Y. V. Grinkova, S. G. Sligar, *Biochem. Biophys. Res. Commun.* **2009**, *387*, 169-173.
- [24] M. Numazawa, M. Nagaoka, N. Sohtome, *Biochemistry* **2005**, *44*, 10839-10845.
- [25] S. J. Skinner, M. Akhtar, *Biochem. J.* **1969**, *114*, 75-81.
- [26] Y. Khatri, A. Luthra, R. Duggal, S. G. Sligar, *FEBS Lett.* **2014**, *588*, 3117-3122.
- [27] J. T. Groves, G. A. McClusky, *J. Am. Chem. Soc.* **1976**, *98*, 859-861.
- [28] P. R. O. de Montellano, *Chem. Rev.* **2010**, *110*, 932-948.
- [29] T. M. Makris, K. von Koenig, I. Schlichting, S. G. Sligar, *J. Inorg. Biochem.* **2006**, *100*, 507-518.
- [30] J. Rittle, M. T. Green, *Science* **2010**, *330*, 933-937.
- [31] I. Schlichting, J. Berendzen, K. Chu, A. M. Stock, S. A. Maves, D. E. Benson, B. M. Sweet, D. Ringe, G. A. Petsko, S. G. Sligar, *Science* **2000**, *287*, 1615-1622.
- [32] S. G. Sligar, *Science* **2010**, *330*, 924-925.
- [33] S. G. Sligar, T. M. Makris, I. G. Denisov, *Biochem. Biophys. Res. Commun.* **2005**, *338*, 346-354.
- [34] M. Newcomb, R. Shen, S.-Y. Choi, P. H. Toy, P. F. Hollenberg, A. D. N. Vaz, M. J. Coon, *J. Am. Chem. Soc.* **2000**, *122*, 2677-2686.
- [35] M. Newcomb, R. Zhang, R. E. P. Chandrasena, J. A. Halgrimson, J. H. Horner, T. M. Makris, S. G. Sligar, *J. Am. Chem. Soc.* **2006**, *128*, 4580-4581.
- [36] H. Isobe, S. Yamanaka, M. Okumura, K. Yamaguchi, J. Shimada, *J. Phys. Chem. B* **2011**, *115*, 10730-10738.
- [37] I. G. Denisov, T. M. Makris, S. G. Sligar, I. Schlichting, *Chem. Rev.* **2005**, *105*, 2253-2277.
- [38] R. E. P. Chandrasena, K. P. Vatsis, M. J. Coon, P. F. Hollenberg, M. Newcomb, *J. Am. Chem. Soc.* **2003**, *126*, 115-126.
- [39] M. Newcomb, D. Aebischer, R. Shen, R. E. P. Chandrasena, P. F. Hollenberg, M. J. Coon, *J. Am. Chem. Soc.* **2003**, *125*, 6064-6065.
- [40] M. C. Gregory, I. G. Denisov, Y. V. Grinkova, Y. Khatri, S. G. Sligar, *J. Am. Chem. Soc.* **2013**, *135*, 16245-16247.
- [41] S. Shaik, S. P. de Visser, F. Ogliaro, H. Schwarz, D. Schroder, *Curr. Opin. Chem. Biol.* **2002**, *6*, 556-567.

- [42] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, K. Schulten, *J. Comput. Chem.* **2005**, *26*, 1781-1802.
- [43] H. Li, A. D. Robertson, J. H. Jensen, *Proteins: Struct., Funct., Bioinf.* **2005**, *61*, 704-721.
- [44] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, *J. Am. Chem. Soc.* **1996**, *118*, 11225-11236.
- [45] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926-935.
- [46] J. Baker, *J. Comput. Chem.* **1997**, *18*, 1079-1095.
- [47] J. Baker, A. Kessi, B. Delley, *J. Chem. Phys.* **1996**, *105*, 192-212.
- [48] R. H. Byrd, P. Lu, J. Nocedal, C. Zhu, *SIAM J. Sci. Comput.* **1995**, *16*, 1190-1208.
- [49] S. Marti, V. Moliner, I. Tunon, *J. Chem. Theory Comput.* **2005**, *1*, 1008-1016.
- [50] Y. K. Zhang, H. Y. Liu, W. T. Yang, *J. Chem. Phys.* **2000**, *112*, 3483-3492.
- [51] B. H. Besler, K. M. Merz, P. A. Kollman, *J. Comput. Chem.* **1990**, *11*, 431-439.
- [52] U. C. Singh, P. A. Kollman, *J. Comput. Chem.* **1984**, *5*, 129-145.
- [53] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648-5652.
- [54] B. Miehllich, A. Savin, H. Stoll, H. Preuss, *Chem. Phys. Lett.* **1989**, *157*, 200-206.
- [55] A. Altun, D. Kumar, F. Neese, W. Thiel, *J. Phys. Chem. A* **2008**, *112*, 12904-12910.
- [56] H. Chen, J. S. Song, W. Z. Lai, W. Wu, S. Shaik, *J. Chem. Theory Comput.* **2010**, *6*, 940-953.
- [57] S. Shaik, D. Kumar, S. P. de Visser, A. Altun, W. Thiel, *Chem. Rev.* **2005**, *105*, 2279-2328.
- [58] P. J. Hay, W. R. Wadt, *J. Chem. Phys.* **1985**, *82*, 299-310.
- [59] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Rob, J. R. Cheeseman, J. A. Montgomery Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, Revision D.02 ed., Gaussian, Inc., Wallingford, CT, **2004**.
- [60] M. J. Field, M. Albe, C. Bret, F. Proust-De Martin, A. Thomas, *J. Comput. Chem.* **2000**, *21*, 1088-1100.
- [61] E. Wigner, *Z. Phys. Chem. B-Chem. E* **1932**, *19*, 203-216.
- [62] R. P. Bell, *The tunnel effect in chemistry*, University Press, Cambridge, **1980**.
- [63] F. Louis, C. A. Gonzalez, J.-P. Sawerysyn, *J. Phys. Chem. A* **2004**, *108*, 10586-10593.
- [64] E. R. Simpson, M. S. Mahendroo, G. D. Means, M. W. Kilgore, M. M. Hinshelwood, S. Grahamlarence, B. Amarneh, Y. J. Ito, C. R. Fisher, M. D. Michael, C. R. Mendelson, S. E. Bulun, *Endocr. Rev.* **1994**, *15*, 342-355.
- [65] I. H. Williams, *J. Mol. Struct.: THEOCHEM* **1983**, *94*, 275-284.
- [66] K. B. Wiberg, *Tetrahedron* **1968**, *24*, 1083-1096.
- [67] Q. Wang, X. Sheng, J. H. Horner, M. Newcomb, *J. Am. Chem. Soc.* **2009**, *131*, 10629-10636.
- [68] S. P. de Visser, F. Ogliaro, P. K. Sharma, S. Shaik, *J. Am. Chem. Soc.* **2002**, *124*, 11809-11826.
- [69] E. J. Klinker, S. Shaik, H. Hirao, L. Que, *Angew. Chem., Int. Ed.* **2009**, *48*, 1291-1295.
- [70] F. Ogliaro, M. Filatov, S. Shaik, *Eur. J. Inorg. Chem.* **2000**, 2455-2458.
- [71] X. Zhang, Y. Liu, Y. Wang, *Theor. Chem. Acc.* **2014**, *133*, 1-8.
- [72] S. P. de Visser, *J. Am. Chem. Soc.* **2006**, *128*, 9813-9824.
- [73] Z. Guo, X. Guan, J.-S. Huang, W.-M. Tsui, Z. Lin, C.-M. Che, *Chem. - Eur. J.* **2013**, *19*, 11320-11331.
- [74] X. Sheng, J. H. Horner, M. Newcomb, *J. Am. Chem. Soc.* **2008**, *130*, 13310-13320.
- [75] P. H. Toy, M. Newcomb, M. J. Coon, A. D. N. Vaz, *J. Am. Chem. Soc.* **1998**, *120*, 9718-9719.
- [76] C. V. Sastri, J. Lee, K. Oh, Y. J. Lee, T. A. Jackson, K. Ray, H. Hirao, W. Shin, J. A. Halfen, J. Kim, L. Que, S. Shaik, W. Nam, *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 19181-19186.
- [77] J. I. Manchester, J. P. Dinnocenzo, Higgins, J. P. Jones, *J. Am. Chem. Soc.* **1997**, *119*, 5069-5070.
- [78] M. Filatov, N. Harris, S. Shaik, *Angew. Chem., Int. Ed.* **1999**, *38*, 3510-3512.
- [79] K. L. Stone, R. K. Behan, M. T. Green, *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 16563-16565.

- 1  
2 [80] D. Fishelovitch, C. Hazan, H. Hirao, H. J. Wolfson, R. Nussinov, S. Shaik, *J. Phys. Chem. B* **2007**,  
3 *111*, 13822-13832.  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review

## Table Captions

**Table 1.** Relative free energies obtained by means of RRHO ( $\Delta G$ ) and FEP ( $\Delta F$ ) for the hydroxylation of ASD in the doublet and quartet spin states (ZPE included in both cases). The basis set used in each calculation is also shown in parentheses. The imaginary frequencies associated with the transition states obtained in the PES explorations are also reported. All energies are expressed in kcal/mol, and frequencies in  $\text{cm}^{-1}$ .

**Table 2.** Primary deuterium ( $^2\text{H}$ ) KIE for the hydrogen transfer process for both gas-phase model and enzymatic system. Corrected KIEs incorporate the tunneling effects by means of the Wigner approximation.

**Table 3.** Atomic Spin Densities of selected atoms or fragments included in the QM atoms of the enzymatic model with the B2 basis set. The values are presented for both spin states: doublet and quartet (in brackets).

**Table 4.** Main geometrical parameters (distances in Å and angles in degrees) found in the active site of the enzyme at the stationary points (see Scheme 1 for atom labeling). The values are presented for both spin states: doublet and quartet (in brackets).

**Table 5.** Relative QM/MM energy decomposition terms (values in kcal/mol) and gas-phase energies ( $\Delta E_{\text{vac}}$ ) for the hydrogen abstraction rate-limiting step. Absolute values with the corresponding errors are shown in brackets. Lennard-Jones term has been broken down into substrate (ASD) and cofactor (Cys+heme).

## Figure Captions

**Scheme 1.** Catalytic process of conversion of ASD into aromatic estrone catalyzed by cytochrome CYP19A aromatase.

**Scheme 2:** Hydroxylation of substrates via cytochromes P450.

**Figure 1.** (a) Atoms comprising QM model in the enzymatic system (Cpd I, ASD substrate and residue Cys-437). (b) Atoms comprising the QM-only model (Cpd I and ASD models, the latter containing only the steroidal rings A and B).

**Figure 2.** Snapshot of the enzymatic active site showing the TS structure for (a) the hydrogen abstraction and (b) the oxygen rebound steps. The most relevant interactions with the propionate chains of the Cpd I, the Cys-437 and the substrate are included.

## Supporting Information

### QM/MM modelling of the hydroxylation of the androstenedione substrate catalyzed by Cytochrome P450 Aromatase (CYP19A1)

*Ignacio Viciano, Raquel Castillo and Sergio Marti\**

Departament de Química Física i Analítica, Universitat Jaume I, 12071 Castelló, Spain.

\* to whom correspondence should be addressed

[smarti@uji.es](mailto:smarti@uji.es)

For Peer Review

**Table S1:** Mulliken Atomic Charges of different atoms or fragments included in the QM atoms of the enzymatic system obtained with the B2 basis set. The values are presented for both spin states: doublet and quartet (in brackets).

Fragment	Hydrogen Abstraction			Oxygen Rebound		
	Reactant	TS (H-Abs)	Intermediate	Intermediate	TS (rebound)	Product
SCH <sub>3</sub> <sup>-</sup>	-0.65 [-0.64]	-0.62 [-0.64]	-0.62 [-0.61]	-0.61 [-0.47]	-0.56 [-0.53]	-0.44 [-0.58]
Fe	1.07 [1.06]	1.06 [1.06]	0.92 [0.92]	0.94 [1.03]	0.97 [1.05]	0.96 [1.07]
O	-0.42 [-0.42]	-0.52 [-0.52]	-0.52 [-0.52]	-0.52 [-0.51]	-0.50 [-0.50]	-0.38 [-0.38]
N <sub>α</sub> +N <sub>β</sub> +N <sub>γ</sub> +N <sub>δ</sub>	-2.05 [-2.05]	-2.00 [-2.00]	-1.99 [-1.98]	-1.99 [-2.19]	-2.04 [-2.19]	-2.27 [-2.38]
Porphyrin <sup>*</sup>	1.88 [1.87]	1.70 [1.73]	1.74 [1.77]	1.76 [1.68]	1.55 [1.64]	1.51 [1.64]
Propionate <sub>δ</sub>	-0.91 [-0.91]	-0.91 [-0.91]	-0.91 [-0.91]	-0.91 [-0.91]	-0.91 [-0.91]	-0.91 [-0.91]
Propionate <sub>α</sub>	-0.89 [-0.89]	-0.90 [-0.90]	-0.90 [-0.90]	-0.89 [-0.89]	-0.90 [-0.89]	-0.90 [-0.89]
C	-0.21 [-0.22]	-0.29 [-0.29]	-0.24 [-0.23]	-0.21 [-0.16]	-0.08 [-0.12]	0.07 [0.07]
H	0.14 [0.15]	0.32 [0.32]	0.31 [0.27]	0.24 [0.26]	0.23 [0.25]	0.27 [0.27]
Substrate <sup>**</sup>	0.05 [0.04]	0.16 [0.15]	0.21 [0.19]	0.18 [0.15]	0.24 [0.19]	0.09 [0.10]

**Table S2:** Main distances (in Å) found in the active site of the aromatase enzyme during both steps, the hydrogen abstraction and the oxygen rebound. Values correspond to both the doublet and the quartet (in brackets) spin states. See Figure 1a for atom labeling.

Distance	Hydrogen Abstraction			Oxygen Rebound		
	Reactant	TS (H-Abs)	Intermediate	Intermediate	TS (rebound)	Product
Fe - N <sub>α</sub>	2.007 [2.006]	2.003 [2.011]	2.018 [2.020]	2.013 [1.984]	1.995 [1.997]	1.990 [1.983]
Fe - N <sub>β</sub>	2.013 [2.017]	2.007 [2.011]	2.013 [2.014]	2.019 [2.002]	2.015 [2.009]	2.001 [2.007]
Fe - N <sub>γ</sub>	2.012 [2.009]	2.026 [2.017]	2.017 [2.017]	2.020 [2.031]	2.021 [2.012]	2.015 [2.005]
Fe - N <sub>δ</sub>	2.015 [2.016]	2.012 [2.016]	2.011 [2.013]	2.024 [2.006]	2.026 [2.006]	2.012 [2.016]
Fe - S	2.681 [2.659]	2.588 [2.564]	2.527 [2.488]	2.520 [2.503]	2.540 [2.574]	2.395 [2.586]
Fe - O	1.625 [1.626]	1.761 [1.761]	1.801 [1.802]	1.806 [1.792]	1.875 [1.831]	2.874 [2.816]
O - C <sub>19</sub>	2.907 [2.906]	2.525 [2.518]	2.895 [2.871]	2.937 [2.960]	2.508 [2.602]	1.423 [1.422]
C <sub>19</sub> - H <sub>19</sub>	1.091 [1.092]	1.399 [1.394]	2.044 [2.182]	2.457 [3.374]	2.500 [2.612]	1.969 [1.964]
O - H <sub>19</sub>	2.487 [2.383]	1.133 [1.134]	0.976 [0.974]	0.973 [0.974]	0.971 [0.974]	0.971 [0.970]
S - H <sub>N,Ala-438</sub>	1.970 [1.958]	2.002 [1.994]	1.982 [1.976]	1.966 [1.975]	1.996 [1.976]	2.007 [1.964]
S - H <sub>N,Gly-439</sub>	2.372 [2.378]	2.366 [2.355]	2.359 [2.361]	2.416 [2.409]	2.425 [2.406]	2.453 [2.429]
S - H <sub>N,Lys-440</sub>	2.733 [2.707]	2.902 [2.935]	2.837 [2.796]	2.732 [2.776]	2.801 [2.764]	2.776 [2.653]
Prop-O <sub>1α</sub> - H <sub>E,Arg-115</sub>	4.999 [5.005]	5.007 [5.007]	5.004 [5.001]	5.050 [5.037]	5.051 [5.036]	5.050 [5.040]
Prop-O <sub>1α</sub> - H <sub>H11,Arg-115</sub>	3.458 [3.457]	3.453 [3.454]	3.452 [3.452]	3.430 [3.432]	3.428 [3.435]	3.430 [3.435]
Prop-O <sub>1α</sub> - H <sub>H12,Arg-115</sub>	1.689	1.682	1.681	1.665	1.662	1.664

	[1.688]	[1.682]	[1.682]	[1.668]	[1.670]	[1.671]
Prop-O <sub>1α</sub> - H <sub>H21,Arg-115</sub>	3.492 [3.497]	3.495 [3.495]	3.495 [3.492]	3.548 [3.550]	3.549 [3.549]	3.548 [3.548]
Prop-O <sub>1α</sub> - H <sub>H22,Arg-115</sub>	1.704 [1.710]	1.705 [1.703]	1.706 [1.705]	1.773 [1.772]	1.777 [1.768]	1.774 [1.771]
Prop-O <sub>1α</sub> - H <sub>E,Arg-375</sub>	1.587 [1.581]	1.584 [1.585]	1.586 [1.579]	1.623 [1.623]	1.617 [1.629]	1.624 [1.630]
Prop-O <sub>1α</sub> - H <sub>H11,Arg-375</sub>	5.123 [5.113]	5.134 [5.135]	5.135 [5.126]	5.187 [5.203]	5.183 [5.211]	5.189 [5.208]
Prop-O <sub>1α</sub> - H <sub>H12,Arg-375</sub>	5.479 [5.471]	5.481 [5.481]	5.483 [5.476]	5.568 [5.587]	5.564 [5.595]	5.569 [5.590]
Prop-O <sub>1α</sub> - H <sub>H21,Arg-375</sub>	4.160 [4.150]	4.170 [4.171]	4.171 [4.165]	4.341 [4.366]	4.342 [4.370]	4.342 [4.367]
Prop-O <sub>1α</sub> - H <sub>H22,Arg-375</sub>	2.719 [2.718]	2.705 [2.705]	2.709 [2.709]	2.806 [2.809]	2.804 [2.812]	2.806 [2.810]
Prop-O <sub>2α</sub> - H <sub>E,Arg-115</sub>	7.162 [7.165]	7.171 [7.171]	7.168 [7.165]	7.236 [7.225]	7.238 [7.225]	7.236 [7.229]
Prop-O <sub>2α</sub> - H <sub>H11,Arg-115</sub>	5.345 [5.335]	5.391 [5.393]	5.383 [5.375]	5.294 [5.328]	5.299 [5.330]	5.301 [5.332]
Prop-O <sub>2α</sub> - H <sub>H12,Arg-115</sub>	3.660 [3.653]	3.695 [3.696]	3.689 [3.682]	3.636 [3.656]	3.639 [3.657]	3.641 [3.660]
Prop-O <sub>2α</sub> - H <sub>H21,Arg-115</sub>	5.439 [5.445]	5.419 [5.418]	5.421 [5.421]	5.534 [5.519]	5.532 [5.521]	5.531 [5.521]
Prop-O <sub>2α</sub> - H <sub>H22,Arg-115</sub>	3.711 [3.715]	3.695 [3.692]	3.698 [3.697]	3.797 [3.786]	3.797 [3.786]	3.796 [3.789]
Prop-O <sub>2α</sub> - H <sub>E,Arg-375</sub>	2.648 [2.644]	2.656 [2.655]	2.655 [2.648]	2.635 [2.637]	2.633 [2.637]	2.638 [2.641]
Prop-O <sub>2α</sub> - H <sub>H11,Arg-375</sub>	5.466 [5.462]	5.477 [5.477]	5.477 [5.471]	5.466 [5.474]	5.459 [5.475]	5.467 [5.474]
Prop-O <sub>2α</sub> - H <sub>H12,Arg-375</sub>	5.110 [5.108]	5.105 [5.105]	5.107 [5.103]	5.125 [5.136]	5.117 [5.138]	5.124 [5.134]
Prop-O <sub>2α</sub> - H <sub>H21,Arg-375</sub>	3.194 [3.191]	3.186 [3.186]	3.190 [3.188]	3.248 [3.252]	3.241 [3.252]	3.243 [3.248]
Prop-O <sub>2α</sub> - H <sub>H22,Arg-375</sub>	1.553 [1.550]	1.549 [1.550]	1.552 [1.548]	1.587 [1.592]	1.579 [1.594]	1.584 [1.590]



Prop-O <sub>1δ</sub> - H <sub>E,Arg-115</sub>	5.293 [5.305]	5.243 [5.228]	5.248 [5.251]	5.399 [5.378]	5.396 [5.376]	5.389 [5.374]
Prop-O <sub>1δ</sub> - H <sub>H11,Arg-115</sub>	7.597 [7.593]	7.572 [7.563]	7.576 [7.576]	7.664 [7.663]	7.666 [7.663]	7.662 [7.661]
Prop-O <sub>1δ</sub> - H <sub>H12,Arg-115</sub>	6.650 [6.640]	6.644 [6.637]	6.645 [6.644]	6.690 [6.705]	6.692 [6.704]	6.689 [6.702]
Prop-O <sub>1δ</sub> - H <sub>H21,Arg-115</sub>	3.598 [3.594]	3.595 [3.585]	3.592 [3.592]	3.643 [3.642]	3.645 [3.638]	3.634 [3.636]
Prop-O <sub>1δ</sub> - H <sub>H22,Arg-115</sub>	4.762 [4.747]	4.785 [4.780]	4.781 [4.778]	4.780 [4.790]	4.782 [4.789]	4.777 [4.785]
Prop-O <sub>1δ</sub> - H <sub>E1,Trp-141</sub>	1.639 [1.631]	1.652 [1.663]	1.650 [1.642]	1.636 [1.645]	1.641 [1.655]	1.664 [1.666]
Prop-O <sub>1δ</sub> - H <sub>E,Arg-145</sub>	4.926 [4.915]	4.929 [4.944]	4.929 [4.928]	5.045 [5.060]	5.048 [5.063]	5.062 [5.068]
Prop-O <sub>1δ</sub> - H <sub>H11,Arg-145</sub>	1.674 [1.670]	1.669 [1.674]	1.672 [1.676]	1.682 [1.677]	1.682 [1.674]	1.685 [1.677]
Prop-O <sub>1δ</sub> - H <sub>H12,Arg-145</sub>	3.045 [3.042]	3.061 [3.061]	3.060 [3.063]	3.044 [3.045]	3.043 [3.039]	3.038 [3.038]
Prop-O <sub>1δ</sub> - H <sub>H21,Arg-145</sub>	5.228 [5.219]	5.223 [5.227]	5.224 [5.228]	5.215 [5.217]	5.214 [5.214]	5.213 [5.213]
Prop-O <sub>1δ</sub> - H <sub>H22,Arg-145</sub>	5.733 [5.724]	5.726 [5.735]	5.728 [5.729]	5.798 [5.802]	5.799 [5.801]	5.806 [5.804]
Prop-O <sub>1δ</sub> - H <sub>E,Arg-435</sub>	2.672 [2.663]	2.646 [2.641]	2.648 [2.645]	2.721 [2.713]	2.722 [2.716]	2.730 [2.720]
Prop-O <sub>1δ</sub> - H <sub>H11,Arg-435</sub>	6.131 [6.128]	6.101 [6.100]	6.104 [6.101]	6.309 [6.306]	6.312 [6.313]	6.324 [6.319]
Prop-O <sub>1δ</sub> - H <sub>H12,Arg-435</sub>	5.727 [5.718]	5.688 [5.690]	5.693 [5.689]	5.845 [5.847]	5.850 [5.855]	5.867 [5.863]
Prop-O <sub>1δ</sub> - H <sub>H21,Arg-435</sub>	5.133 [5.140]	5.128 [5.124]	5.127 [5.124]	5.304 [5.301]	5.304 [5.305]	5.308 [5.307]
Prop-O <sub>1δ</sub> - H <sub>H22,Arg-435</sub>	3.665 [3.675]	3.664 [3.656]	3.662 [3.659]	3.821 [3.814]	3.820 [3.816]	3.818 [3.816]
Prop-O <sub>2δ</sub> - H <sub>E,Arg-115</sub>	3.076 [3.091]	3.031 [3.014]	3.037 [3.040]	3.181 [3.164]	3.177 [3.162]	3.170 [3.157]
Prop-O <sub>2δ</sub> - H <sub>H11,Arg-115</sub>	5.652	5.624	5.631	5.699	5.696	5.691

	[5.661]	[5.615]	[5.636]	[5.698]	[5.697]	[5.689]
Prop-O <sub>2δ</sub> - H <sub>H12,Arg-115</sub>	4.988 [4.994]	4.973 [4.969]	4.976 [4.982]	5.004 [5.015]	5.002 [5.012]	4.998 [5.006]
Prop-O <sub>2δ</sub> - H <sub>H21,Arg-115</sub>	1.621 [1.623]	1.620 [1.618]	1.619 [1.623]	1.638 [1.640]	1.640 [1.635]	1.631 [1.632]
Prop-O <sub>2δ</sub> - H <sub>H22,Arg-115</sub>	3.275 [3.273]	3.289 [3.292]	3.287 [3.290]	3.260 [3.268]	3.260 [3.266]	3.256 [3.260]
Prop-O <sub>2δ</sub> - H <sub>E1,Trp-141</sub>	3.123 [3.122]	3.138 [3.145]	3.140 [3.139]	3.139 [3.147]	3.136 [3.150]	3.149 [3.151]
Prop-O <sub>2δ</sub> - H <sub>E,Arg-145</sub>	6.961 [6.959]	6.967 [6.983]	6.967 [6.968]	7.081 [7.097]	7.083 [7.098]	7.092 [7.100]
Prop-O <sub>2δ</sub> - H <sub>H11,Arg-145</sub>	3.609 [3.606]	3.593 [3.607]	3.596 [3.600]	3.604 [3.595]	3.607 [3.594]	3.614 [3.602]
Prop-O <sub>2δ</sub> - H <sub>H12,Arg-145</sub>	4.249 [4.248]	4.262 [4.279]	4.259 [4.260]	4.217 [4.221]	4.222 [4.219]	4.226 [4.226]
Prop-O <sub>2δ</sub> - H <sub>H21,Arg-145</sub>	6.618 [6.610]	6.611 [6.628]	6.610 [6.614]	6.589 [6.592]	6.593 [6.591]	6.597 [6.596]
Prop-O <sub>2δ</sub> - H <sub>H22,Arg-145</sub>	7.410 [7.408]	7.404 [7.421]	7.405 [7.408]	7.459 [7.462]	7.461 [7.461]	7.469 [7.465]
Prop-O <sub>2δ</sub> - H <sub>E,Arg-435</sub>	1.589 [1.599]	1.574 [1.572]	1.582 [1.586]	1.616 [1.621]	1.607 [1.622]	1.618 [1.615]
Prop-O <sub>2δ</sub> - H <sub>H11,Arg-435</sub>	5.294 [5.303]	5.293 [5.282]	5.299 [5.299]	5.370 [5.377]	5.361 [5.380]	5.367 [5.371]
Prop-O <sub>2δ</sub> - H <sub>H12,Arg-435</sub>	5.155 [5.164]	5.143 [5.138]	5.151 [5.154]	5.195 [5.205]	5.187 [5.208]	5.195 [5.199]
Prop-O <sub>2δ</sub> - H <sub>H21,Arg-435</sub>	4.057 [4.064]	4.090 [4.080]	4.089 [4.084]	4.148 [4.159]	4.142 [4.162]	4.144 [4.155]
Prop-O <sub>2δ</sub> - H <sub>H22,Arg-435</sub>	2.376 [2.382]	2.421 [2.410]	2.417 [2.410]	2.476 [2.486]	2.471 [2.490]	2.471 [2.483]
O' - H <sub>D2,Asp-309</sub>	1.646 [1.665]	1.695 [1.692]	1.668 [1.667]	3.157 [3.185]	3.176 [3.207]	3.128 [3.149]
O'' - H <sub>N,Met-374</sub>	1.768 [1.767]	1.823 [1.821]	1.787 [1.793]	1.784 [1.771]	1.808 [1.794]	1.824 [1.815]
N <sub>α</sub> - H <sub>19</sub>	3.085 [3.021]	2.870 [2.857]	2.549 [2.464]	2.490 [3.370]	2.648 [2.811]	2.782 [2.807]

$N_6 - H_{19}$	4.295	3.408	3.171	2.793	2.416	2.603
	[4.198]	[3.382]	[2.926]	[2.347]	[2.381]	[2.543]

For Peer Review

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Table S3:** Main geometrical parameters (distances in Å and angles in degrees) obtained in the QM-only model during the hydrogen abstraction step for both doublet and quartet (in brackets) spin states. See Figure 1b for atom labeling.

	Hydrogen Abstraction		
	Reactant	TS (H-Abs)	Intermediate
Fe - S	2.639 [2.636]	2.542 [2.462]	2.482 [2.412]
Fe - O	1.625 [1.626]	1.752 [1.748]	1.807 [1.799]
O -C <sub>19</sub>	3.510 [3.532]	2.532 [2.532]	3.372 [3.358]
C <sub>19</sub> -H <sub>19</sub>	1.094 [1.094]	1.409 [1.413]	3.274 [3.288]
O -H <sub>19</sub>	2.494 [2.506]	1.141 [1.143]	0.973 [0.974]
S-Fe-O	169.9 [171.3]	174.0 [174.7]	171.3 [171.7]
C <sub>19</sub> -H <sub>19</sub> -O	153.9 [155.8]	166.4 [164.1]	87.3 [85.7]
H <sub>19</sub> -O-Fe-N <sub>α</sub>	80.0 [79.5]	45.0 [53.8]	43.7 [74.6]

**Table S4.** Relative zero point energy (ZPE) and Gibbs contributions to the free energy (in kcal/mol) for the hydroxylation of ASD in the doublet and quartet spin states, using the B1 basis set. The Gibbs contribution includes both the ZPE and the thermal corrections, based on the rigid-rotor and the harmonic-oscillator (RRHO) approximations.

	ZPE	Gibbs
$^2\text{R}$	0.0	0.0
$^4\text{R}$	0.0	-0.4
$^2\text{TS}_{\text{H-abs}}$	-5.8	-5.6
$^4\text{TS}_{\text{H-abs}}$	-6.1	-5.4
$^2\text{I}$	-3.3	-3.1
$^4\text{I}$	-2.9	-4.0
$^2\text{TS}_{\text{reb}}$	-5.3	-2.9
$^4\text{TS}_{\text{reb}}$	-3.5	-3.0
$^2\text{P}$	-0.7	1.3
$^4\text{P}$	-0.9	1.3

### Details for the FEP error calculation:

The uncertainties reported for FEP energies in present work are based on the first-order expansion of the free energy, expressed as  $\pm 2\sigma$ :

$$\Delta F_{i \rightarrow j} = -RT \ln \left( \frac{1}{N} \sum_k e^{-\frac{\Delta U_k}{RT}} \right) \pm RT \frac{2\delta}{\frac{1}{N} \sum_k e^{-\frac{\Delta U_k}{RT}}}$$

$$\delta^2 = \frac{1+2\tau}{N} \left[ \left( \frac{1}{N} \sum_k e^{-2\frac{\Delta U_k}{RT}} \right) - \left( \frac{1}{N} \sum_k e^{-\frac{\Delta U_k}{RT}} \right)^2 \right]$$

where  $1+2\tau$  is the sampling ratio, defined as:

$$1+2\tau = \frac{1+r_1}{1-r_1} \quad r_1 = \frac{\sum_{i=2}^N (x_i - \bar{x})(x_{i-1} - \bar{x})}{\sum (x_i - \bar{x})^2}$$

For further details, please consider the following references:

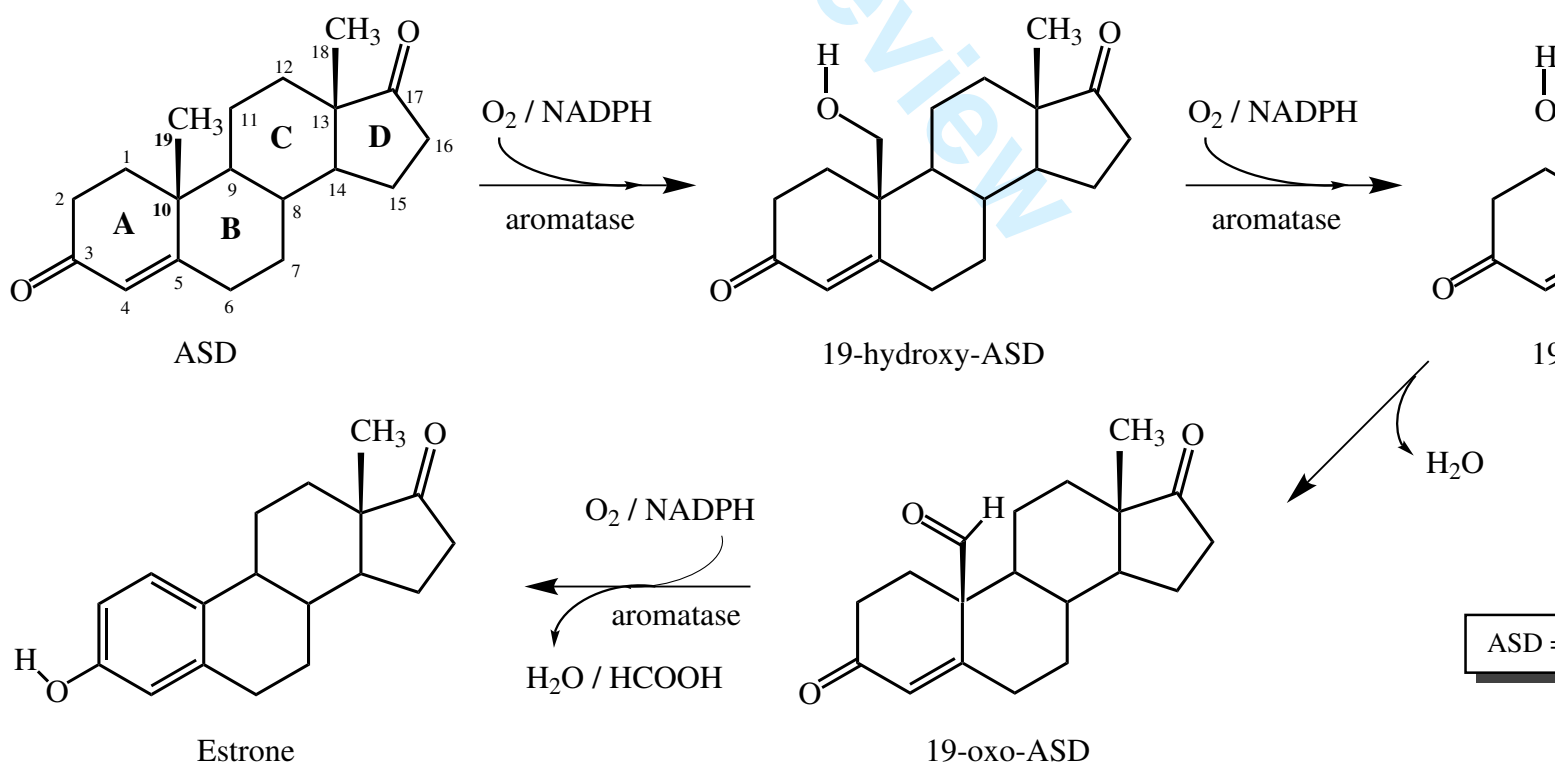
[1] Chipot, C.

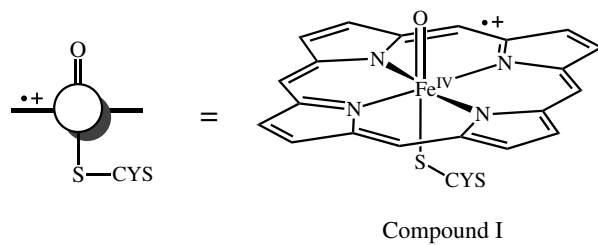
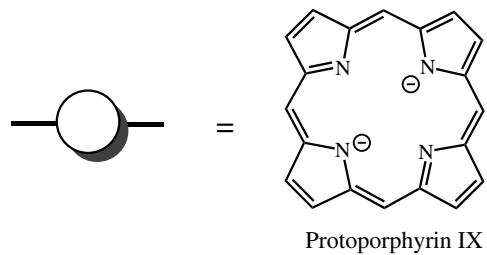
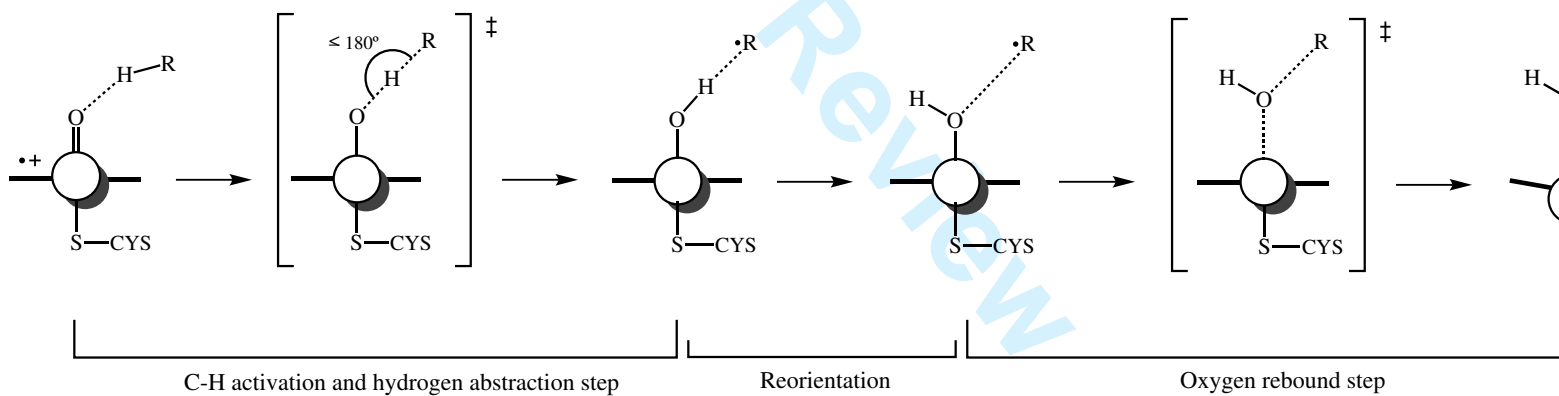
Free Energy Calculations in Biological Systems. How Useful Are They in Practice?

*New Algorithms for Macromolecular Simulation, Springer Berlin Heidelberg, 2006, 49, 185-211*

[2] T. P. Straatsma, H. J. C. Berendsen, A. J. Stam, *Molecular Physics* **1986**, 57, 89-95

[3] E. B. Smith, B. H. Wells, *Molecular Physics* **1984**, 52, 701-704





R = S



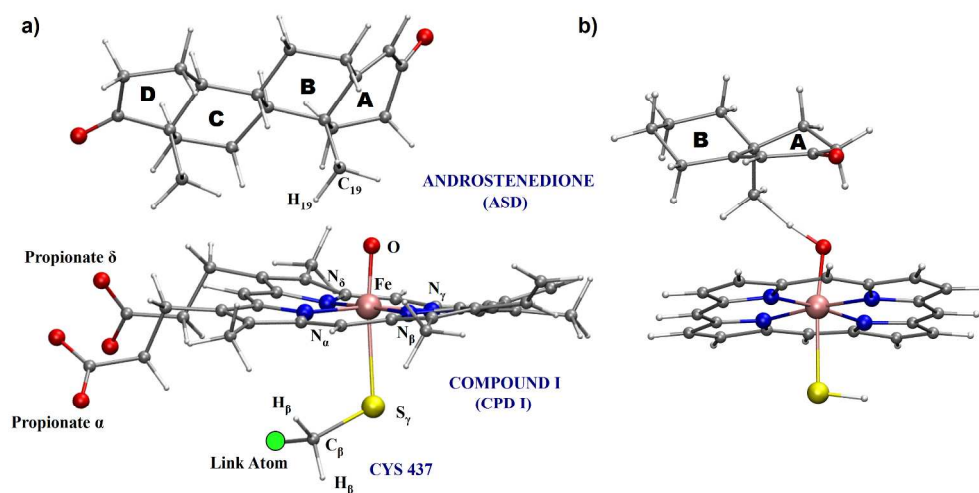


Figure 1. (a) Atoms comprising QM model in the enzymatic system (Cpd I, ASD substrate and residue Cys-437). (b) Atoms comprising the QM-only model (Cpd I and ASD models, the latter containing only the steroidal rings A and B).  
1247x634mm (72 x 72 DPI)

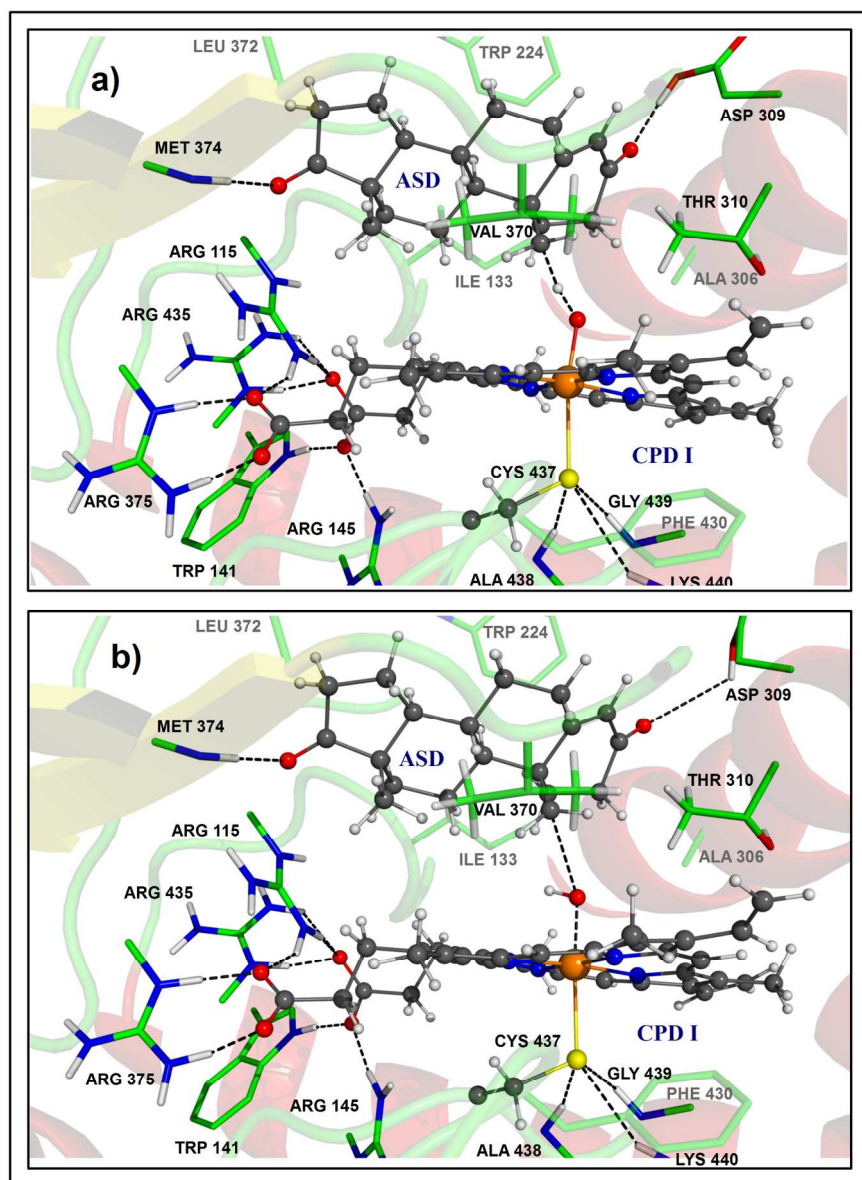


Figure 2. Snapshot of the enzymatic active site showing the TS structure for (a) the hydrogen abstraction and (b) the oxygen rebound steps. The most relevant interactions with the propionate chains of the Cpd I, the Cys-437 and the substrate are included.  
651x880mm (72 x 72 DPI)

**Table 1.** Relative free energies obtained by means of RRHO ( $\Delta G$ ) and FEP ( $\Delta F$ ) for the hydroxylation of ASD in the doublet and quartet spin states (ZPE included in both cases). The basis set used in each calculation is also shown in parentheses. The imaginary frequencies associated with the transition states obtained in the PES explorations are also reported. All energies are expressed in kcal/mol, and frequencies in  $\text{cm}^{-1}$ .

	$\Delta G_{\text{vac}}$ (B1)	$\Delta G_{\text{enz}}$ (B1)	$\Delta F_{\text{enz}}$ (B1)	$\Delta G_{\text{enz}}$ (B2)	$\Delta F_{\text{enz}}$ (B2)
$^2\text{R}$	0.0	0.0	0.0	0.0	0.0
$^4\text{R}$	-0.1	1.4	2.9 $\pm$ 0.1	1.9	3.0
$^2\text{TS}_{\text{H-abs}}$	25.7 (1601 <i>i</i> )	22.3 (1606 <i>i</i> )	13.5 $\pm$ 0.4	20.0 (1619 <i>i</i> )	11.4
$^4\text{TS}_{\text{H-abs}}$	25.8 (1713 <i>i</i> )	24.5 (1645 <i>i</i> )	13.5 $\pm$ 0.3	21.8 (1675 <i>i</i> )	12.0
$^2\text{I}$	16.3	17.1	11.5 $\pm$ 0.6	13.5	8.7
$^4\text{I}$	16.4	17.2	12.3 $\pm$ 0.4	14.7	9.3
$^2\text{TS}_{\text{reb}}$		20.0 (258 <i>i</i> )	13.0 $\pm$ 0.8	15.5 (230 <i>i</i> )	10.0
$^4\text{TS}_{\text{reb}}$		19.3 (174 <i>i</i> )	15.1 $\pm$ 1.0	17.0 (119 <i>i</i> )	12.6
$^2\text{P}$		-32.4		-36.9	
$^4\text{P}$		-42.7		-43.3	

**Table 2.** Primary deuterium ( $^2\text{H}$ ) KIE for the hydrogen transfer process for both gas-phase model and enzymatic system. Corrected KIEs incorporate the tunneling effects by means of the Wigner approximation.

Spin state	Gas-Phase model		Enzyme Aromatase	
	$^2\text{H}_{19}$ primary KIE	$^2\text{H}_{19}$ corrected KIE	$^2\text{H}_{19}$ primary KIE	$^2\text{H}_{19}$ corrected KIE
doublet	7.027	10.225	7.281±0.004	10.315±0.006
quartet	7.013	10.445	7.074±0.003	10.456±0.007

For Peer Review

**Table 3.** Atomic Spin Densities of selected atoms or fragments included in the QM atoms of the enzymatic model with the B2 basis set. The values are presented for both spin states: doublet and quartet (in brackets).

Fragment	Hydrogen Abstraction			Oxygen Rebound		
	Reactant	TS (H-Abs)	Intermediate	Intermediate	TS (rebound)	Product
SCH <sub>3</sub> <sup>-</sup>	-0.20 [0.17]	-0.21 [0.14]	-0.14 [0.10]	-0.10 [-0.08]	-0.19 [-0.06]	-0.17 [0.28]
Fe	1.30 [1.22]	1.06 [0.96]	1.02 [0.92]	1.02 [2.06]	1.16 [2.24]	1.34 [2.70]
O	0.83 [0.83]	0.49 [0.52]	0.13 [0.15]	0.09 [0.13]	-0.06 [0.04]	-0.00 [0.01]
Porphyrin	-0.93 [0.78]	-0.87 [0.82]	-0.95 [0.87]	-0.98 [-0.09]	-0.62 [-0.10]	-0.17 [0.01]
C <sub>19</sub>	0.00 [0.00]	0.60 [0.62]	1.03 [1.04]	1.04 [1.05]	0.76 [0.94]	0.00 [0.00]
H <sub>19</sub>	-0.00 [-0.00]	-0.05 [-0.05]	-0.01 [0.00]	0.00 [0.01]	0.00 [0.02]	0.00 [0.00]
Substrate*	0.00 [0.00]	-0.02 [-0.02]	-0.07 [-0.07]	-0.07 [-0.07]	-0.05 [-0.06]	0.00 [0.00]

\* The Substrate does not include the C<sub>19</sub> and H<sub>19</sub> atoms

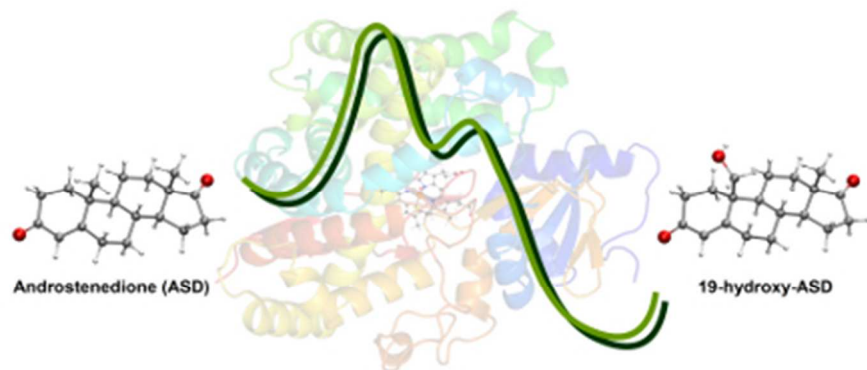
**Table 4.** Main geometrical parameters (distances in Å and angles in degrees) found in the active site of the enzyme at the stationary points (see Scheme 1 for atom labeling). The values are presented for both spin states: doublet and quartet (in brackets).

	Hydrogen Abstraction step			Oxygen Rebound step		
	Reactant	TS (H-Abs)	Intermediate	Intermediate	TS (rebound)	Product
Fe-S	2.681 [2.659]	2.588 [2.564]	2.527 [2.488]	2.520 [2.503]	2.540 [2.574]	2.395 [2.586]
Fe-O	1.625 [1.626]	1.761 [1.761]	1.801 [1.802]	1.806 [1.792]	1.875 [1.831]	2.874 [2.816]
O-C <sub>19</sub>	2.907 [2.906]	2.525 [2.518]	2.895 [2.871]	2.937 [2.960]	2.508 [2.602]	1.423 [1.422]
C <sub>19</sub> -H <sub>19</sub>	1.091 [1.092]	1.399 [1.394]	2.044 [2.182]	2.457 [3.374]	2.500 [2.612]	1.969 [1.964]
O-H <sub>19</sub>	2.487 [2.383]	1.133 [1.134]	0.976 [0.974]	0.973 [0.974]	0.971 [0.974]	0.971 [0.970]
O'-H <sub>D2,Asp-309</sub>	1.646 [1.665]	1.695 [1.692]	1.668 [1.667]	3.157 [3.185]	3.176 [3.207]	3.128 [3.149]
O''-H <sub>N,Met-374</sub>	1.768 [1.767]	1.823 [1.821]	1.787 [1.793]	1.784 [1.771]	1.808 [1.794]	1.824 [1.815]
S-Fe-O	173.9 [173.6]	170.2 [170.2]	172.4 [172.8]	173.1 [171.9]	176.2 [172.8]	171.3 [171.6]
C <sub>19</sub> -H <sub>19</sub> -O	101.5 [107.6]	171.2 [169.6]	144.5 [126.6]	110.1 [57.1]	79.3 [78.7]	43.1 [43.2]
H <sub>19</sub> -O-Fe-N <sub>α</sub>	12.3 [11.6]	23.1 [20.2]	17.5 [-5.1]	-17.7 [-111.8]	-49.0 [-64.7]	-57.6 [-60.2]

**Table 5.** Relative QM/MM energy decomposition terms (values in kcal/mol) and gas-phase energies ( $\Delta E_{\text{vac}}$ ) for the hydrogen abstraction rate-limiting step. Absolute values with the corresponding errors are shown in brackets. Lennard-Jones term has been broken down into substrate (ASD) and cofactor (Cys+heme).

	$\Delta E_{\text{elec}}$		$E_{\text{pol}}$		$E_{\text{LJ}}$		$\Delta E_{\text{LJ}} (\text{ASD})$	$\Delta E_{\text{LJ}} (\text{Cys+heme})$	$\Delta E_{\text{vac}}$
<sup>2</sup> R	[-630.4±7.9]	0.0	[79.0±1.5]	0.0	[-316.9±7.1]	0.0	0.0	0.0	0.0
<sup>2</sup> TS	[-629.0±8.7]	1.4	[76.1±2.3]	-2.9	[-321.0±9.3]	-4.1	-2.6	-1.5	25.4
<sup>4</sup> R	[-621.4±6.9]	9.0	[80.4±1.8]	1.4	[-324.4±7.8]	-7.5	0.7	-8.2	0.17
<sup>4</sup> TS	[-620.4±8.0]	10.0	[74.5±1.5]	-4.5	[-328.9±8.6]	-12.0	-4.2	-7.8	26.3

For Peer Review



155x67mm (72 x 72 DPI)

Peer Review

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



1 CYP19A1 aromatase is a Cytochrome P450 responsible for the final step of the androgens  
2 conversion into the corresponding estrogens, and thus can play a significant role in the  
3 hormone-dependent breast cancer development. The first of three oxidative subcycles of  
4 this enzyme consists of an initial hydrogen abstraction followed by a oxygen rebound  
5 step, giving place to an hydroxylated form of the androstenedione substrate via two  
6 possible electronic spin pathways.  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review