The Mechanism of Human Aromatase (CYP 19A1) Revisited: DFT and QM/MM Calculations Support a Compound I-Mediated Pathway for the Aromatization Process

Balázs Krámos, Julianna Oláh[*]

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

ABSTRACT: Human aromatase is responsible for the last step of estrogen biosynthesis, for the aromatization of ring A of androstenedione (ASD) or testosterone. In this work the mechanism of aromatization was studied using gas phase and hybrid QM/MM calculations. It is shown that human aromatase can efficiently catalyze the aromatization process via a Compound I (or Compound II) mediated pathway. The nature of the oxidant is very sensitive to the polarizing environment of the enzyme, as the oxidant has a Compound I nature in the gas phase

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

e-mail: julianna.olah@mail.bme.hu; telephone number: +36-1-463-1286

calculations, which is modulated by the enzyme environment to become a mixed Compound I and Compound II character. The electronic structure of the obtained QM-only and QM/MM stationary points is thoroughly discussed.

1. INTRODUCTION

Cytochrome P450 enzymes (CYPs) form the superfamily of thiolate-ligated heme enzymes. Although the functioning of CYPs is very versatile, most members are monooxygenases.[1] This means that these enzymes have a unique ability to activate molecular oxygen at ambient temperatures by binding O_2, breaking the O-O double bond.[2] One of the oxygen atoms is reduced to water, while the other one is used to the formation of a highly reactive species, called Compound I,[3] which is able to oxidize a large variety of substrates. As such, CYPs are in the frontline of protecting the human body against xenobiotics by oxidizing them to more soluble and more easily excretable compounds. CYPs are also responsible for the metabolism of 90% of drugs taken by men, thereby influence their efficacy that has to be considered during drug discovery and development. Despite the crucial role of CYPs in the pharmaceutical industry, the issue of their regio- and stereoselectivity is still hampering. The majority of oxidizing CYPs are promiscuous, most of them metabolize several substrates, and from a single substrate various products may be obtained.

An exception is human aromatase that acts upon only two, highly similar substrates (testosterone and androstenedione) and catalyzes their aromatization to estrogens without forming any known by-products (apart from the intermediates of the reaction). Aromatization is the last step, and the bottle–neck reaction of estrogen biosynthesis in which androstenedione is converted to estrone (see **Scheme 1**) or testosterone is converted to estradiol. Furthermore, aromatase also influences the steroid levels in the body.[4] As elevated estrogen levels are known to increase significantly the risk of the development and reappearance of hormone-dependent cancers, e.g. breast cancer, inhibitors of aromatase are frequently administered during their treatment.[5,6,7] The development of more specific and more potent inhibitors could be facilitated

by a detailed knowledge of the enzyme's functioning, still there are many debates regarding the catalytic mechanism of aromatase.

Scheme 1. Conversion of androstenedione (ASD) to estrone.
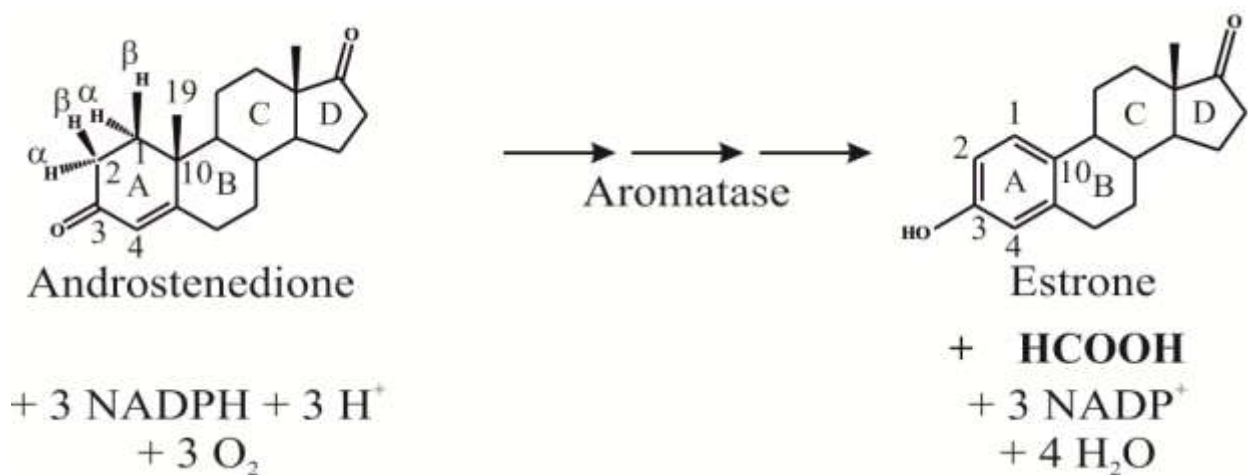
The overall catalytic cycle of aromatase consists of three sub-cycles, each requiring 1 mol of NADPH and 1 mol of O_2.[8,9] It is commonly accepted that in the first and second sub-cycles the classical CYP-mediated hydrogen atom abstraction-hydroxyl radical rebound mechanism[10] is responsible for the hydroxylation of the C_{19}-methyl group, leading to the formation of a C_{19}-*gem*-diol, which may undergo dehydration to yield the C_{19}-aldehyde.[11,12] In contrast to the first two sub-cycles, the reaction mechanism of the third sub-cycle has been much debated e.g mechanisms including 2β-hydroxylation,[13,14,15] 4,5-epoxidation,[16,17] Baeyer-Villiger oxidation of C_{19} and 10β-hydroxyestr-4-ene-3,17-dione formation[20] were suggested, but which were rejected by later experimental works. It is known that in the third sub-cycle the oxidative cleavage of the C_{10}-C_{19} bond occurs and the aromatized steroid and one molecule of formic acid are generated. KIE measurements prove that the 1β- and 2β-hydrogen atoms are eliminated selectively,[18,19] and ^{18}O labeling studies indicate that the third oxygen atom is incorporated into formate.[9,20] For decades the most accepted hypothesis has been the formation of a peroxohemiacetal after the

nucleophilic attack of the ferrous superoxo complex on the C_{19} carbon (see **Scheme 2.** for the suggested reaction mechanism and for its structure).[9,20] We note that the ferrous superoxo complex, has been traditionally called the ferric peroxo complex, but based on its electronic structure – one unpaired electron located on the O-O moiety, and six electrons in the d-block of iron[30] – the name ferrous superoxo complex is more realistic.

Scheme 2. Estrone generation via peroxohemiacetal formation from enolized (red) or non-enolized 19-oxo-ASD in the last catalytic sub-cycle of aromatase.
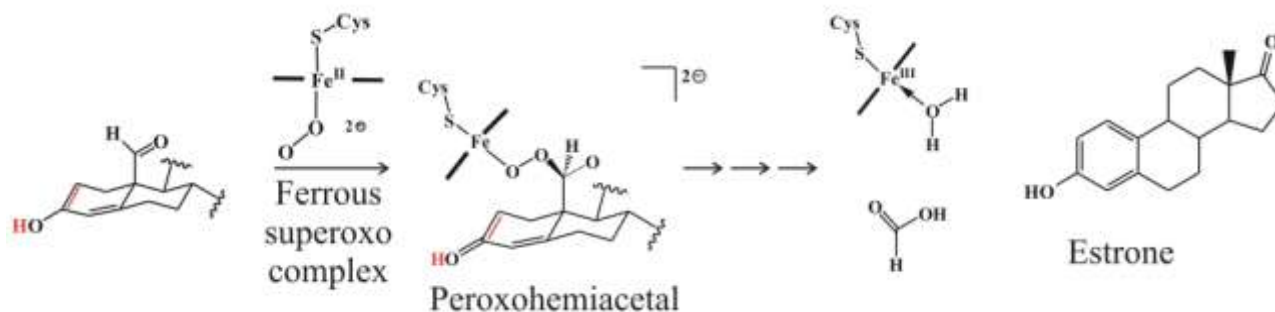
The mechanism of peroxohemiacetal formation is supported indirectly by an EPR study, in which the ferrous superoxo complex was identified at 77 K after the cryogenic reduction of aromatase, indicating its increased stability[21] compared to P450$_{cam}$[22] whose Compound 0 state was obtained in a similar study. Based on these experiments it can be hypothesized that the proton delivery pathway in aromatase is hindered leading to the increased stability of the ferrous superoxo complex, which facilitates peroxohemiacetal formation. This is in sharp contrast to the consensus catalytic cycle of most CYPs, in which fast proton delivery to the ferrous superoxo complex leads to the formation of Compound 0 followed by Compound I generation (see **Scheme 3**), which carries out the actual oxidation reaction.

Scheme 3. Schematic overview of the transformation of the ferrous superoxo complex to Compound I according to the consensus catalytic cycle of CYPs. Compound II is the one electron reduced form of Compound I.
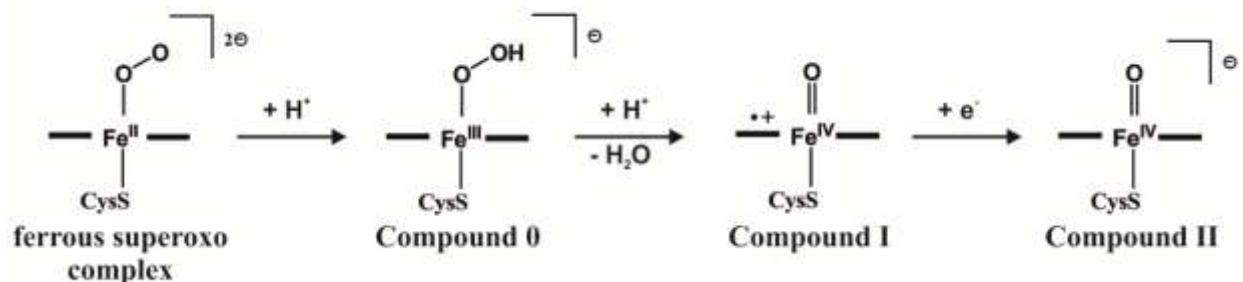
Based on a homology model of aromatase a hydrogen bond between the formyl group of the ligand and the Thr_{310} side chain was suggested to interrupt the proton delivery pathway and help the formation of peroxohemiacetal.[23] The presence of such a hydrogen bond is supported by resonance Raman measurements which show that if carbon monoxide binds to the reduced enzyme ligated to 19-oxo-ASD, the environment of carbon monoxide is less polar than in the case of the enzyme ligated to 19-nor-ASD (19-oxo-ASD without the C_{19}-formyl group) ligand.[24] Based on gas-phase quantum mechanical[25] and combined quantum mechanics molecular mechanics (QM/MM) methods[26] Hackett at al suggested that the peroxohemiacetal intermediate is formed in a barrierless process in aromatase, and their proposed mechanism is in accordance with experimental findings on the selective elimination of the 1β- and 2β-hydrogens.[18,19] Although these results may support the formation of a peroxohemiacetal intermediate, its existence is not directly proven. Furthermore, a very recent resonance Raman study found no significant difference between the oxy-complexes of aromatase in the first and third sub-cycles implicating the involvement of Compound I in the terminal lyase step.[27] Kinetic solvent isotope effect studies of the three sub-cycles of aromatase led to an identical conclusion.[28]

Both experimental and theoretical studies clearly indicate that the reactivity of CYPs is not a clear-cut mechanism. In many cases these enzymes are characterized by multistate reactivity[29,30] implying that different spin states and various reaction routes close in energy contribute to catalysis and from a given substrate various products might be obtained. An interesting example is the CYP125 enzyme, which converts cholest-4-en-3-one to cholest-4-en-3-one-26-acid in the course of sequential oxidation of the terminal methyl group of the side chain of cholest-4-en-3-one to the alcohol, aldehyde and finally carboxylic acid. However, five additional metabolites of the enzyme were also identified.[31] While the mechanism leading to the aldehyde intermediate seems to follow the consensus catalytic cycle of most oxidizing CYPs,[29,30] involving Compound I, it seems likely that in the last sub-cycle various reaction routes could be followed. On the major pathway, leading to the carboxylic acid, oxidation of the formyl group by Compound I has been suggested[32] while the identified by-products are supposed to derive from the deformylation of this aldehyde via peroxohemiacetal formation.[31]

This example calls the attention to the competition between the various reaction routes available in CYPs. QM/MM offer the possibility gain atomistic insight into the reactivity of enzymes, it allows us to test various hypotheses regarding the reaction mechanism, and agreement with experiment can be used to validate the results. In QM/MM calculations the atoms directly involved in the chemical reaction are treated quantum mechanically, while the rest of the system is treated at a lower level, using molecular mechanics. The available methods differ in various aspects,[33] e.g. which QM level and MM force field they use, how the QM and MM regions are connected (mechanical vs electrostatic embedding), how they treat the QM-MM boundary (e.g. using link atoms[34,35] or frozen orbitals[36,37,38,39]), in the methodology of QM/MM optimization (whether to use microiterations[40] or not), if they allow the determination of only

reaction energies (e.g adiabatic mapping) or of free energies (e.g. umbrella sampling[41] or free energy perturbation) as well. Until very recently, QM/MM free energy calculations could only be performed using a semi-empirical Hamiltonian, but advances in both computing facilities and in methodologies begin to allow QM/MM free energy calculations using ab initio or density functional methods (examples include GTPases[42] and phosphate hydrolasis in water[43]). Importantly, metadynamics,[44] a free energy sampling technique and algorithms such as paradynamics[45] which is an efficient tool for perturbatively estimating the high-level of theory correction on the reaction paths in condensed phases obtained with a different potential, have contributed significantly to this field, and there is hope that soon these calculations become available for cytochrome P450 enzymes as well.

In the present work we use a combination of molecular dynamics simulations and QM/MM adiabatic mapping to investigate the recently proposed hypothesis that Compound I catalyses the terminal lyase step in aromatase. Based on our previous work,[46] selective 2β-hydrogen elimination[18] from ASD and the enolization of substrate molecule could be coupled to the protonation of the ferrous superoxo complex in an energetically favorable process leading to Compound 0 (see (a) in **Scheme 4**), which could be easily converted to Compound I via proton coupled electron transfer (PCET).[47] Therefore, in this study we investigated the likeliness of various reaction routes in which Compound I carries out the last step of aromatization, similarly to the majority of CYP-catalyzed monooxygengase reactions.

We only studied those possible reaction routes which are in accordance with the experimentally known properties of the aromatization reaction: (1) the 1β-hydrogen is eliminated enzymatically,[48] (2) C_{10}-C_{19} bond scission occurs in the last catalytic sub-cylce, (3) one oxygen

atom derived from the oxygen molecule required for the third catalytic sub-cycle is built into the produced formic acid molecule.[9,49]

The studied pathways suiting these requirements are depicted in **Scheme 4**. The first step in each case is selective 1β-hydrogen abstraction from the enolized 19-oxo-ASD by Compound I and the formation of an intermediate (**B**). The multi-state reactivity (MSR) of CYPs allows both radical (**B$_{rad}$**) and cationic (**B$_{cat}$**) configurations for the electronic structure of the intermediate **B**. In this work we studied the reactions in the quartet and doublet states, as in several cases the quartet and doublet states of CYPs were shown to catalyze different conversions. E.g. the epoxidation reaction catalyzed by the enzyme on the doublet surface leads to a single product, while a mixture of products might be obtained in quartet state.[50,51,52] The prominent feature of CYPs that the properties of the oxidizing species are very sensitive to interactions with the protein pocket and to external conditions has been called the chameleon effect.[53,54] Starting from intermediate **B** various scenarios exist for the completion of the reaction depending on the site of hydroxyl rebound which can be either C_{19} or C_1. If hydroxyl rebound occurs to C_{19} the process is accompanied by the cleavage of the C_{10}-C_{19} bond, the release of formic acid and the aromatization of ring A. If hydroxyl rebound occurs to C_1 a potential by-product, 1β-hydroxy-19-oxo-ASD (**C**) could be formed which may rearrange to yield **D**. We also investigated a third possibility, which was first suggested (and immediately rejected) by Hackett et al[25]: instead of hydroxyl rebound, the hydroxo complex might abstract another hydrogen atom from the formyl group, which would lead to the production of carbon monoxide and water besides the aromatic steroid (**E**).

Scheme 4. Suggested mechanism for aromatization of 19-oxo-ASD. (a) Selective 2β-hydrogen elimination from ASD leads to the enolized ligand and Compound 0 as suggested by QM/MM calculations.[46] (b) Compound I abstracts the 1β-hydrogen from the ligand leading via various pathways to the final products.
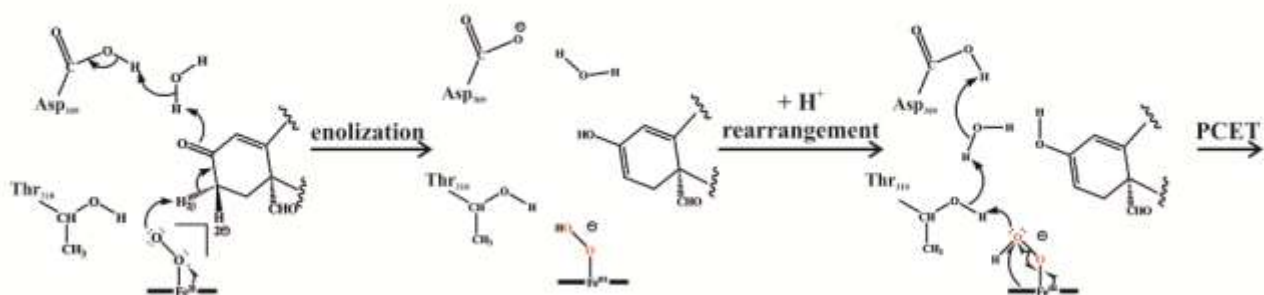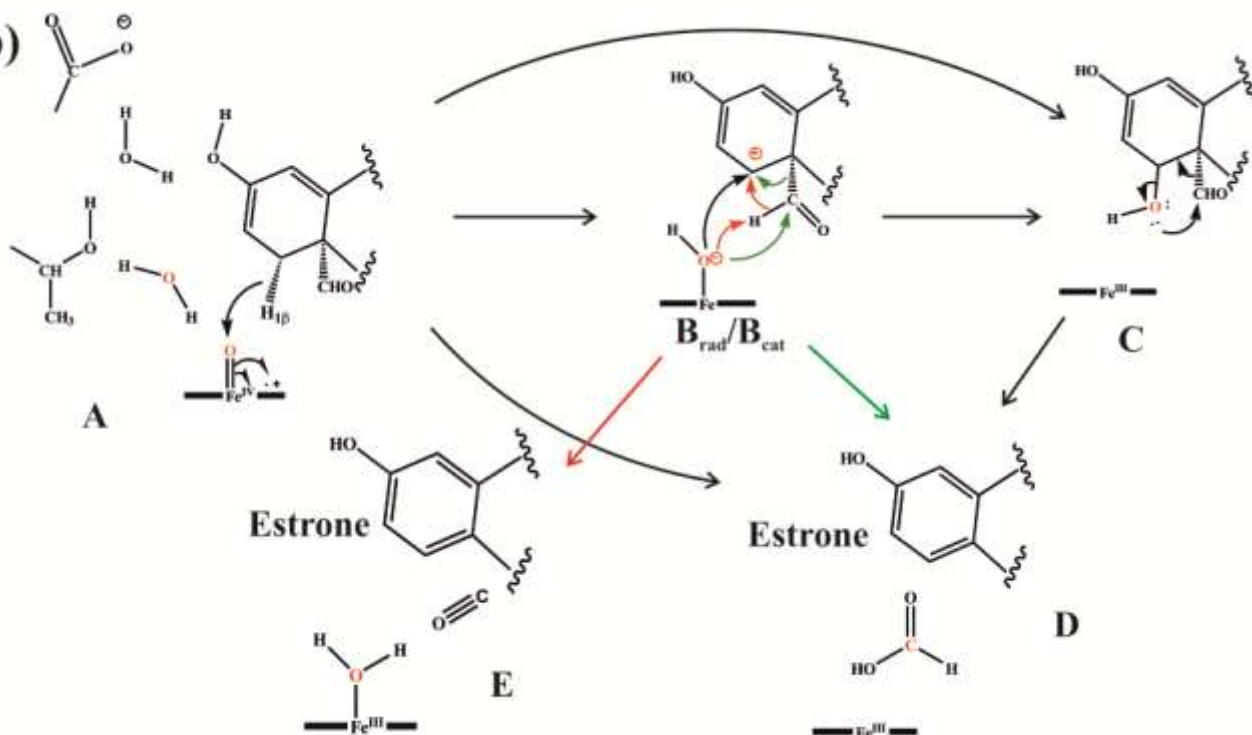
2. METHODS

2.1 QM calculations. In the first approximation the substrate molecule was modelled as ring A and B of the sterane skeleton, and the enzyme by using a simple model of Compound I. In the QM-only models an unsubstituted porphin ring was used instead of the porphyrin IX ring, as it was shown to be a suitable model of it.[55] The axial Cys_{437} ligand was represented by an SH^- group, which performs better in vacuum calculations than the bigger SMe^- group due to fortunate cancellation of errors.[56,57] The calculations were performed with the Gaussian 09 program package[58] with the B3LYP hybrid functional[59] using an unrestricted formalism, which has become the general method for the modeling of the active site of CYPs, and is known to give correct spin state energetics for the studied species.[29] In the two lowest energy states of Compound I, two unpaired electrons are localized in the Fe-O moiety on the π^*_{xz} and π^*_{yz} orbitals, which can be ferromagnetically (quartet) or antiferromagnetically (doublet) coupled to the third unpaired electron found on a porphyrin orbital having a_{2u} symmetry. As these two spin states are almost degenerate,[29,30] the calculations were performed in both the quartet and doublet spin states.

While describing the quartet state of Compound I is straightforward using unrestricted KS theory, the doublet state can only be studied using a broken-symmetry solution. This doublet state has a multireference character, in which the most important component is the $D_1 = |{\uparrow}{\uparrow}{\downarrow}\rangle$ configuration.

In the case of such system the expectancy value of $\langle S^2 \rangle$ is 1.75 in the doublet state due to the UKS treatment, as the highest occupied beta orbital has no occupied pair among the alfa orbitals):

$$\langle S^2 \rangle_{UKS} = \langle S^2 \rangle_{exact} + N_\beta - \sum_{i,j}^{N_{MO}} \left\langle \phi_i^\alpha \middle| \phi_j^\beta \right\rangle^2$$

Although, 1.75 as the expectancy value of $\langle S^2 \rangle$ is not realistic and is not in accordance with the multi-reference calculations, the B3LYP functional is known to give correct energy values for relative energies of different spin states and excitation energies (expect for the A_{2u}-A_{1u} excitation in the doublet Compound I) of the most interesting states in the chemistry of CYPs, except for the splitting of the lowest $A_{1u}-A_{2u}$ radical states in Compound I.[60]

For geometry optimizations the SDD basis set and SDD effective core potential[61,62] were used for Fe in conjunction with the 6-31G* basis set for other atoms, similarly to other studies.[46,63] This basis set composition will be called B1. The identity of the located stationary points were checked using second derivative calculations, minima were characterized by all real vibrational frequencies, and transition states by a single imaginary frequency. Zero-point energy correction was also calculated at this level. The stability of all wavefunctions was checked by generating singly excited determinants, and the wavefunction was considered stable if the corresponding energy eigenvalues of the singly excited determinants were higher than those of the initial wavefunction.

In order to obtain more reliable energies single-point calculations were carried out on the obtained geometries at the B3LYP/cc-pVTZ level (this basis set will be called B2 throughout the figures) and empirical dispersion correction was calculated with the DFT-D3 program[64] developed by Grimme and co-workers. Natural charges and spin populations were calculated with the NBO 5.9 program.[65] In order to check the dependence of the result on the used DFT functional single point calculations were carried out with M06[66], TPSSh[67], OLYP[68] and the long-range and dispersion corrected ωB97X-D[69] functionals in conjunction with the cc-pVTZ basis

set. The effect of solvation on the relative energies of the obtained species was estimated by performing continuum solvent model calculations[70,71] with a dielectric constant ε(water)=78.3553 on the obtained stationary points assuming that the geometries are similar in the gas-phase and in solution.

2.2 System Setup: Molecular Mechanics (MM) calculations and Molecular Dynamics (MD) Simulations. The crystal structure of human aromatase in complex with androstenedione (PDB ID: 3EQM),[72] was used as a starting structure. In the present work the same enzymatic system was studied as in our previous work. Therefore, details of system setup can be found in ref. 46 and in the Supporting Information. Briefly, the protein was solvated in a 25 Å-sphere around the iron center and after an initial optimization of the structure using molecular mechanics a 2 ns long Langevin dynamics simulation was carried out on the system. Before the MD simulation was carried out a single water molecule was placed between the Thr_{310} side chain and the Fe-O moiety of the Compound I, which represented the catalytic water molecule formed in the course of Compound I formation. This water molecule has been previously shown to lower by 4 kcal/mol the energy barriers for hydrogen abstraction in $P450_{cam}$ by increasing the electron affinity of Compound I via hydrogen bonding to the Fe-O moiety.[73,74] In the present study the side chain of Asp_{309} was deprotonated, because this residue serves as an indirect proton source for the conversion of the ferrous superoxo complex to Compound 0 (see **Scheme 4** above) based on our previous work.[46] We supposed that the following reactions of Compound 0 and Compound I will occur faster than Asp_{309} could be re-protonated. The role of the protonation state of Asp_{309} on the reaction mechanism was investigated by QM/MM single point calculations including ionized or neutral sidechains of $Asp_{309.}$ These calculations proved that the protonation

state of Asp$_{309}$ has an almost negligible effect on the reaction mechanism, thus modelling the Asp$_{309}$ sidechain as ionized is acceptable.

Topology file of 19-oxo-ASD and non-standard parameters are given in the SI, while the MM parameters for Compound I are available in the literature.[75] For all simulations the CHARMM27 force field[76] and the CHARMM software package[77] was used. No unrealistic distortion of the active site of the enzyme was observed in the course of the MD simulation.

2.3 QM/MM calculations. Three starting structures were selected from the last nanosecond of the 2 ns long MD trajectory for the QM/MM calculations. The QM region contained the unsubstituted porphine ring, the whole substrate, the side chains of Thr$_{310}$ and Cys$_{437}$ residues and the catalytic water molecule between the Thr$_{310}$ residue and the Fe-O moiety (see **Figure 1** a)). The Thr$_{310}$ and Cys$_{437}$ residues were represented by their side chains broken between their C$_{\alpha}$ and C$_{\beta}$ atoms. Hydrogen-type link atoms were used to cap the open valences of QM atoms. The charges of MM atoms (and the groups they belong to) bonded to QM atoms were set to zero to avoid unphysical effects due to the strong polarization of the QM wave function due to the proximity of large point charges. The atoms further than 20 Å from the iron center were fixed. The geometry of the QM region was optimized at the B3LYP/B1 level and single-point calculations were carried out on the minima and transition state structures at the B3LYP/6-311+G* level. The MM region was described with the CHARMM27 force field.[76] For these simulations the QoMMMa 8.02 program[78] was used which uses the Gaussian 09 program[58] for describing QM region and the TINKER program[79,80] for the MM region. As the inclusion of dispersion effect was shown to improve the accuracy of reaction barriers for CYP catalyzed reactions dispersion effects are important in enzymatic reactions,[81,82] empirical dispersion correction of energy and forces were applied in each optimization steps using DFT-D2[83]

parameters as implemented in the QoMMMa 8.02 program. Three parallel energy profiles were generated by adiabatic mapping using a step size of 0.1 Å. Definition of the reaction coordinates used for the various conversion steps are shown in **Figure 1**. Minima were optimized without constrains and the maxima of the energy profiles were considered as transition states. Careful examination of the reaction profiles shows that this method introduces an uncertainty of about 1 kcal/mol into value of the obtained barrier heights in most cases, which is much smaller than accuracy of the B3LYP method. NBO charges and spin populations were calculated with the NBO5.9 program.[65] Single-point calculations with the hybrid M06,[66] TPSSh,[67] and pure OLYP[68] functionals and the long-range and dispersion corrected ωB97X-D[69] DFT functionals were carried out in conjunction with the 6-311+G* basis set on the QM region. The MOLDEN[84] and VMD[85] programs were used for visualization.
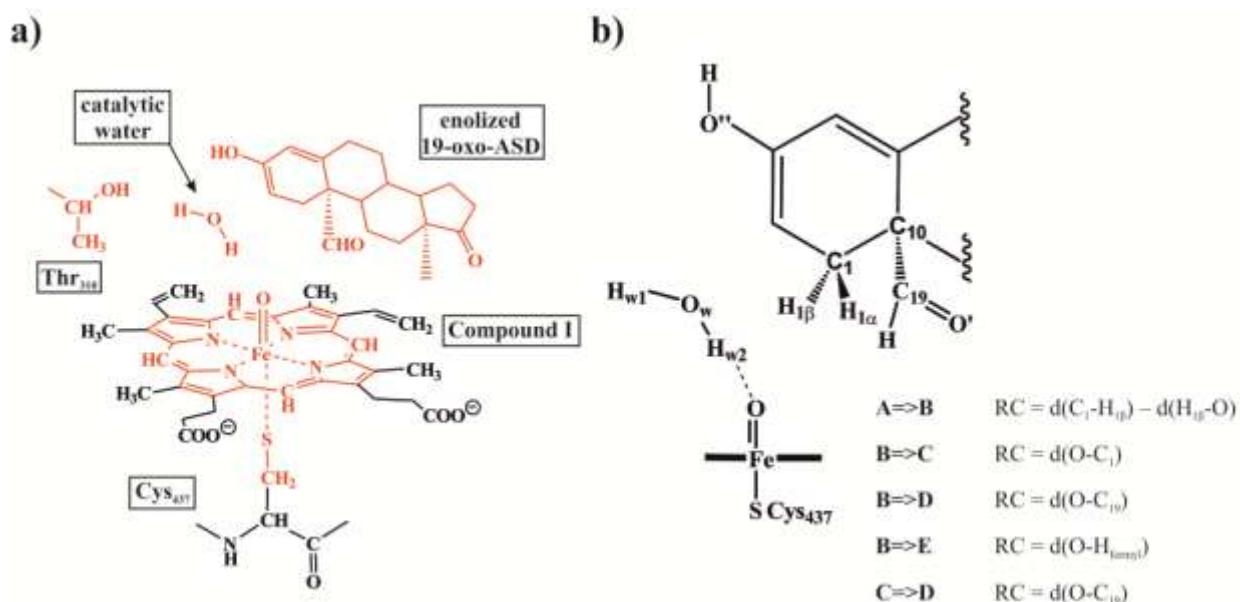

Figure 1. a) Atoms shown in red (together with their bonded hydrogen atoms) constituted the QM region in the QM/MM calculations b) reaction coordinates (RC) applied in the course of adiabatic mapping of the various transformation steps.

3. RESULTS AND DISCUSSION

3.1 QM calculations

3.1.1. The energy landscape. First, gas-phase calculations were carried out on a simplified model of the enzyme-substrate complex. The geometries and the relative energies are shown in **Figure 2** and **Figure 3**. The most important geometrical parameters are summarized in the SI (**Table S1**).

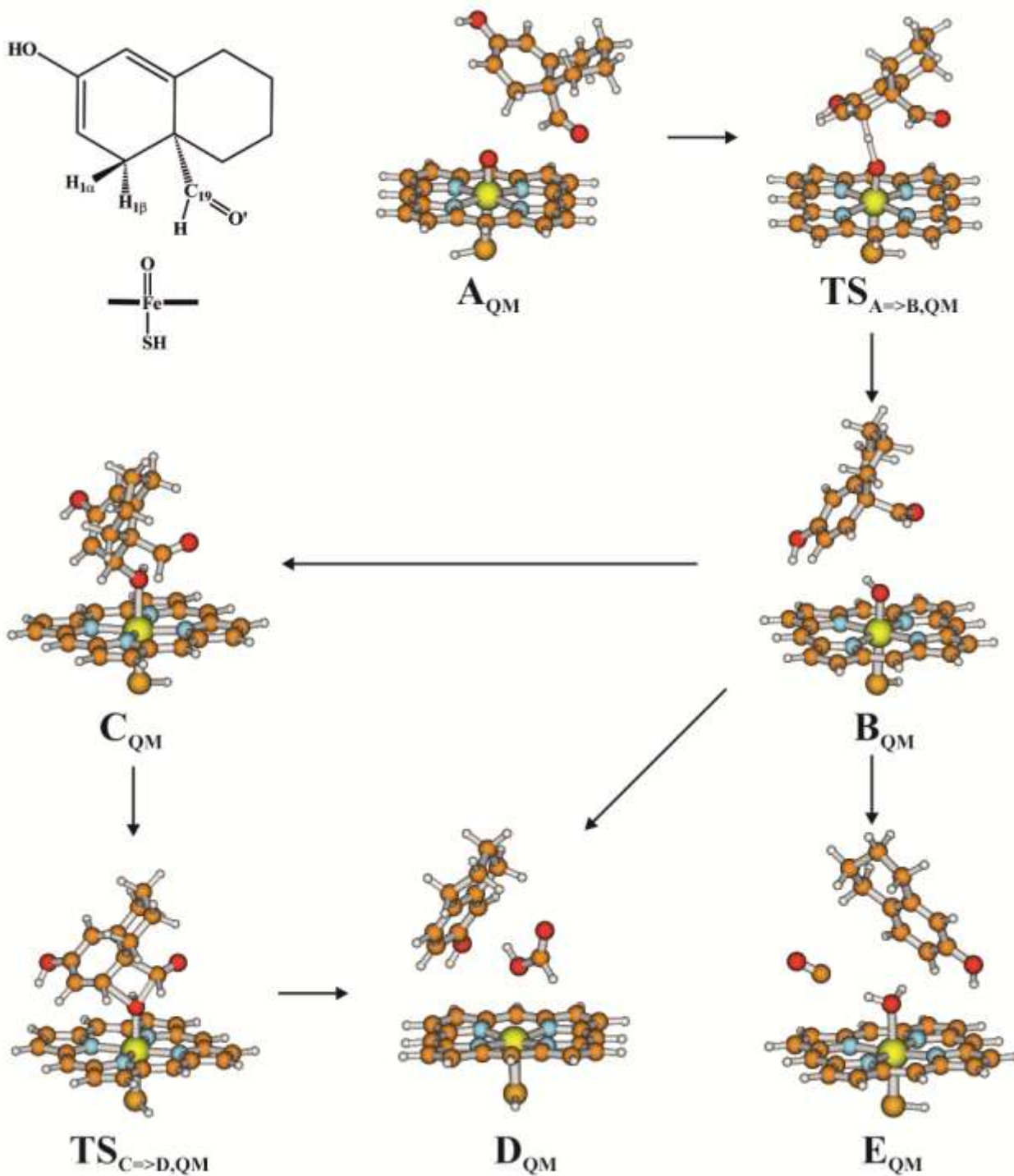**Figure 2.** QM-only model systems. The depicted geometries were obtained in the doublet state, and they are very similar to the corresponding quartet structures except for the **C_QM** and **TS_{C=>D,QM}** models whose Fe-O distances are longer.
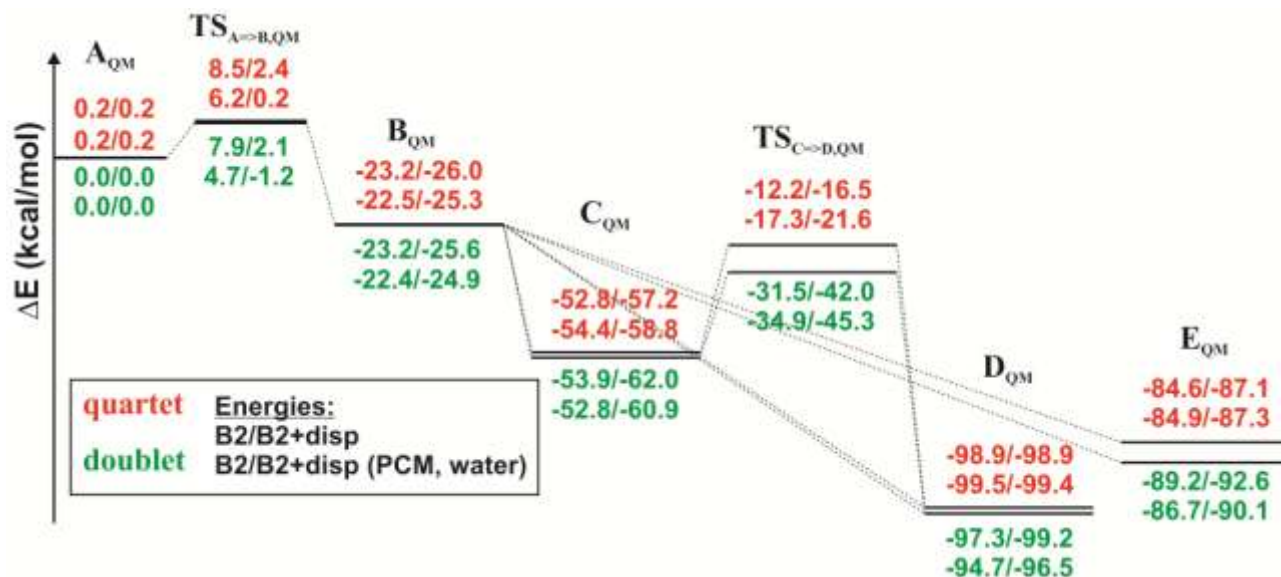
Figure 3. Energy diagram for the investigated reaction routes in the gas-phase and in PCM in the quartet (red) and doublet (green) states using basis set B2 (see label at the bottom of the figure). The PCM results are single-point energies on the gas-phase geometries.

The deformylation reaction is energetically a highly favorable process. The overall reaction heat is close to -100 kcal/mol when estrone and formic acid (D_{QM}) are formed, which is much higher than the reaction heat calculated for other reactions carried out by Compound I. E.g. camphor hydroxylation was predicted by gas-phase calculations to have a reaction heat of about -45 kcal/mol.[86] The much higher reaction heat found in this study is due to the very energy-rich character of the dienol substrate. A large part of the energy associated to aromatization of ring A of the substrate is already released in the first part of the reaction as shown by the low relative energy of the transition state $TS_{A=>B,QM}$ and the intermediate (B_{QM}). The barrier of the initial 1β-hydrogen abstraction step is surprisingly low: the barrier is about 8 kcal/mol in both spin states, which is reduced by about 2 kcal/mol when a continuum solvent model is included in the calculations. The barrier almost diminishes if dispersion effects are taken into account. This is in

18

contrast to the barrier of 19 kcal/mol calculated in P450$_{cam}$ in the gas-phase. The intermediate of aromatase (**B$_{QM}$**) is suggested to be lower in energy by about 23 kcal/mol than the reactants, in contrast to the relative energy of the intermediate of the P450$_{cam}$–hydroxylation reaction (+14 kcal/mol compared to the reactants). The low energy of the intermediate is due to the emerging aromaticity of the ring, as it is also shown by its almost planar structure.

Based on frequency calculations the **B$_{QM}$** state is a real local minimum on the potential energy surface, and it serves as a branching point towards the formation of various products. The natural product of the enzyme (**D$_{QM}$**) is the lowest in energy, while the non-aromatic **C$_{QM}$** is the least stable. We could not locate first order-saddle points for the rebound reaction of the OH radical to the ligand (**B$_{QM}$**=>**C$_{QM}$**, **B$_{QM}$**=>**D$_{QM}$** and **B$_{QM}$**=>**E$_{QM}$** conversions) and relaxed energy scan calculations also showed practically barrierless processes. Previous gas-phase studies has also found that the rebound step of hydroxylation reactions is either barrierless or involves low barriers.[86] This implies that within the accuracy of the presented gas-phase results, it is computationally impossible to rank the likelihood of the investigated reaction channels and to predict the experimental product ratio. Furthermore, most likely the enzyme can drive the catalytic reaction towards its final products via subtle ways, e.g. electrostatic and steric interactions, which cannot be captured by gas-phase calculations.

The pathways leading to **D$_{QM}$** and **E$_{QM}$** both produce the aromatized steroids, thus they could only be distinguished by measuring CO and formic acid production. Based on the experiments only formic acid is formed in the reaction, but small amounts of CO might have remained unnoticed in these experiments. Furthermore, small amounts of carbon monoxide might play a role in regulating the enzyme's activity (autoinhibition). Interestingly, it was shown that the aromatization of 19-nor-ASD (ASD lacking the 19-methyl group, thus from which CO cannot be

produced) was highly inhibited by the atmosphere containing 44% carbon monoxide, in contrast to the aromatization reaction of ASD, 19-hydroxy-ASD and the 19-oxo-ASD from which in principle CO could be produced.[87]

The formation of C_{QM} is much more intriguing. The barrier that needs to be overcome during the rearrangement reaction to yield D_{QM} is about 20-30 kcal/mol (due to the formation of a highly strained four-membered ring structure in the transition state), which might suggest that the 1β-hydroxylated by-product should be observable, still it has not been reported in the case of ASD yet, to the best of our knowledge. However, when aromatase was used to aromatize an artificial substrate, 19-nor-ASD the 1β-hydroxylated by-product was detected in addition to estrone,[88] showing the competition between the two pathways. The most likely reason why the 1β-hydroxylated by-product has not been detected for ASD yet, is that it decomposes much faster to yield the aromatized steroid and formic acid then the 1β-hydroxylated by-product of 19-nor-ASD, which makes it relatively difficult to distinguish this reaction route from the direct pathway. Indeed, in the ^{18}O-labeling study of the reaction mechanism, direct rearrangement of C to yield D was suggested as a possibility.[20] An indication of the easier decomposition of the 1β-hydroxylated derivative of 19-oxo-ASD than that of 19-nor-ASD is that in vacuo the barrier for formic acid elimination is 13-15 kcal/mol lower than the barrier of water elimination from 19-nor-ASD (see **Table S2**), and the reaction is more exothermic by about 20 kcal/mol in the case of 19-oxo-ASD. Alternatively, non-enzymatic decomposition reaction routes might also exist for the decomposition of the 1β-hydroxylated by-product. For instance, the 2β-hydroxy derivative is known to undergo very facile aromatization in aqueous solution at pH 7 with a half-life of 90 sec to yield the aromatic product and formic acid.[13] This reaction was suggested to proceed via a base-catalyzed pathway, which could also facilitate the aromatization of the 1β-hydroxylated by-

product (**Scheme 5**) and alleviate the stress associated to the formation of a four-membered ring during intramolecular direct decomposition. In this case the oxygen atom derived from the oxygen molecule required for the catalytic cycle may not be incorporated into the formate, which is experimentally suggested. However, this is not in direct contrast with the ^{18}O-labeling study[20] as in all experiments signals of unlabeled HCOOH were detected beside the HC[18]OOH signal, with relative intensities up to 30% in the mass spectrum, which might have originated from the non-enzymatic elimination.

Scheme 5. Hypothetic base-catalyzed non-enzymatic aromatization of 1β-hydroxy-19-oxo-ASD.
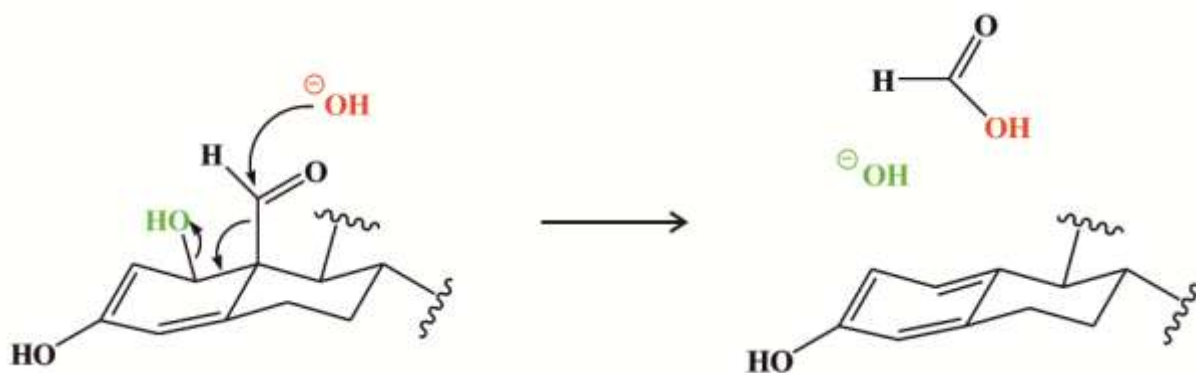
The shape of the calculated reaction profiles and the structure of the located stationary points are rather similar in the two spin states, with the exception of the transition state connecting **C** and **D** (**TS**$_{C=>D,QM}$). While the quartet and doublet states are very close in energy in the case of most structures, the quartet **TS**$_{C=>D,QM}$ state is higher in energy than the doublet by 10-20 kcal/mol depending on applied functional and basis set. This difference is due to the weaker ligand binding in the quartet state, which is most pronounced in the **C**$_{QM}$ and **TS**$_{C=>D,QM}$ states. The Fe-O distances are 2.213 Å, 2.008 Å and 3.294 Å in the doublet **C**$_{QM}$, **TS**$_{C=>D,QM}$ and **D**$_{QM}$

systems and 3.430 Å, 3.437 Å and 3.645 Å in quartet states, which suggest much weaker ligand

binding and a practically dissociated complex in the quartet state. We calculated the energy

barrier for the elimination of formic acid from the ligand of C_{QM} (in the absence of the heme

group) in a unimolecular process in vacuo and obtained a value of 29.4 kcal/mol at the

B3LYP/cc-pVTZ level including dispersion correction. These results suggest that the

[Fe(III)(Porph)SH] complex lowers the barrier for the conversion of **C** to **D** in doublet state, but

in the quartet state the barrier is 10 kcal/mol higher than in a unimolecular process without the

heme complex.

To examine the necessity of the enolization of the ligand for the aromatization process, we

determined the energy barriers for 1β-hydrogen abstraction from the not-enolized 19-oxo-ASD

ligand in both spin states. They turned out to be considerably, by about 14-16 kcal/mol, higher

(see **AK$_{QM}$, TS$_{AK\Rightarrow BK,QM}$,** and **BK$_{QM}$** states in the **Table 1** and in the SI in **Table S1**) in

accordance with other studies emphasizing the essential role of the enolization process.[25]

Table 1. Effect of the DFT functional on the relative energies (in kcal/mol) of the QM-only

model systems calculated at the B3LYP/B1 geometries using the cc-pVTZ basis set

model	multiplicity	B3LYP^a	M06	TPSSh	OLYP	ωB97X-D
A$_{QM}$	doublet	0.0	0.0	0.0	0.0	0.0
	quartet	0.2	0.7	0.4	1.9	0.1
TS$_{A\Rightarrow B,QM}$	doublet	2.1	3.4	3.7	4.3	8.5
	quartet	2.4	4.6	5.2	8.9	8.1
B$_{QM}$	doublet	-25.6	-29.9	-22.7	-21.7	-25.2
	quartet	-26.0	-30.3	-22.2	-19.2	-25.3
C$_{QM}$	doublet	-62.0	-66.5	-50.0	-34.9	-72.0
	quartet	-57.2	-67.5	-42.1	-37.9	-65.3
TS$_{C\Rightarrow D,QM}$	doublet	-42.0	-42.3	-30.9	-15.6	-44.5
	quartet	-16.5	-26.0	-6.3	-15.8	-18.5
D$_{QM}$	doublet	-99.2	-97.1	-89.7	-88.6	-103.7
	quartet	-98.9	-105.2	-86.3	-89.5	-103.3
E$_{QM}$	doublet	-92.6	-92.5	-79.6	-73.5	-97.7

	quartet	-87.1	-93.2	-68.5	-69.5	-90.0
AK$_{QM}$	doublet	-18.6	-18.3	-16.8	-14.8	-18.9
	quartet	-18.3	-17.3	-16.4	-12.7	-18.6
TS$_{AK=>BK,QM}$	doublet	-1.1	-3.5	3.2	6.9	0.6
	quartet	-0.5	-0.9	3.6	9.7	1.4
BK$_{QM}$	doublet	-9.8	-10.1	-10.2	-7.3	-13.3
	quartet	-17.3	-22.9	-15.3	-11.6	-17.7

[a] B3LYP energies are dispersion corrected.

In order to increase the reliability of the obtained data we carried out single-point calculations using various functionals (see **Table 1**). The observed trends calculated with the B3LYP functional are repeated by the other functionals, but the relative energies differ considerably. Results obtained by the M06 functional are the most similar to those of B3LYP, and the OLYP values are the most different. The barrier of 1β-hydrogen abstraction was predicted by all studied functionals to be lower than 10 kcal/mol (and in most cases lower than 5 kcal/mol), confirming our previous conclusion that this is most likely a very fast process. The barrier of the **C$_{QM}$** => **D$_{QM}$** process is predicted to be about 20 kcal/mol by (B3LYP, TPSSh, OLYP), or even larger by M06 and ωB97X-D in the doublet state, and above 35 kcal/mol in the quartet state by most functionals (with exception of OLYP: (22.1 kcal/mol)) supporting the conclusion based on B3LYP that direct rearrangement of **C** to yield **D** is unlikely.

3.1.2. Electronic structure of the obtained stationary points. In order to get insight into the electronic structure of the studied complexes, natural charges and spin populations of selected atoms and fragments have been investigated; a compilation of the data at the B3LYP/cc-pVTZ level is shown in **Figure 4** and in **Table S3** in the SI. Data obtained with other functionals repeat the trends obtained by B3LYP.
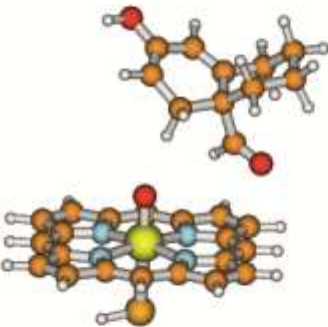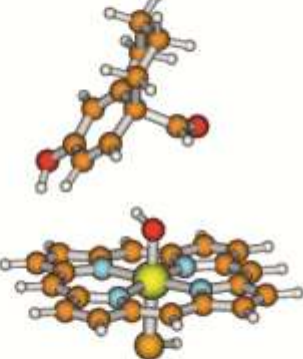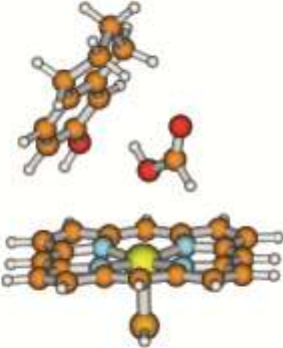
		A$_{QM}$		B$_{QM}$		D$_{QM}$	
		Q	SP	Q	SP	Q	SP
Fe	vacuum	0.98/0.97	1.14/1.06	1.06/1.06	1.65/1.67	0.98/1.24	1.06/2.35
	PCM (water)	1.00/0.99	1.21/1.12	1.08/1.08	1.71/1.74	1.00/1.28	1.08/2.44
O	vacuum	-0.46/-0.46	0.89/0.94	-0.82/-0.82	0.26/0.25	-0.68/-0.68	0.00/0.00
	PCM (water)	-0.54/-0.53	0.83/0.88	-0.85/-0.86	0.23/0.23	-0.68/-0.68	0.00/0.00

Porphyrin	vacuum	-0.37/-0.35	-0.45/0.45	-0.55/-0.56	-0.04/-0.03	-0.81/-0.83	-0.05/0.14
	PCM (water)	-0.13/-0.11	-0.67/0.66	-0.49/-0.50	-0.02/-0.02	-0.76/-0.77	-0.05/0.16
SH$^-$	vacuum	-0.15/-0.16	-0.57/0.54	-0.15/-0.16	0.11/0.11	-0.18/-0.40	-0.01/0.50
	PCM (water)	-0.34/-0.35	-0.37/0.34	-0.21/-0.22	0.06/0.07	-0.25/-0.51	-0.02/0.40
Substrate	vacuum	0.01/0.01	0.00/0.01	0.00/0.01	-0.99/0.98	0.00/0.00	0.00/0.00
	PCM (water)	0.01/0.01	-0.01/0.01	0.00/0.02	-0.99/0.97	0.00/0.00	0.00/0.00

Figure 4. Natural charges (Q in e) and spin populations (SP) for selected atoms and fragments in QM-only calculations. Values belonging to the doublet/quartet states are derived from single point calculations at the B3LYP/cc-pVTZ level. Color coding: yellow: iron; greenish-blue: nitrogen; red: oxygen; brown: carbon; orange: sulfur; white: hydrogen.

The spin populations on the SH$^-$ and porphyrin ligands in the **A$_{QM}$** system suggests one unpaired electron on the a$_{2u}$ orbital in accordance with the electronic configuration of Compound I (see **Figure 5**), although the relative contribution of the porphyrin ring and the SH$^-$ group to this orbital is significantly affected by the using a PCM model in the calculations. Therefore, as expected, the **A$_{QM}$** state is best described as a combination of Compound I and the substrate molecule (both in vacuo and in PCM), which is converted to an uncharged radical (**B$_{QM}$** with ^{2,4}**I$_{rad}$(IV)** character, see **Figure 5**) with -0.99 (doublet state) or 0.98 (quartet state) spin

population in the course of 1β-hydrogen abstraction, and solvation effects do not alter the radical character of the intermediate. Simultaneously, the orbital primarily composed of the a_{2u} porphyrin orbital and the sulfur p_z orbital becomes doubly occupied. The single unpaired electron in the doublet C_{QM}, D_{QM} and E_{QM} states is located on one of the d_{xz} and d_{yz} orbitals of the ferric ion. In contrast, in the quartet state there are three unpaired electrons on the d_{xz}, d_{yz} and σ^*_{Fe-S} (containing d_{z2} orbital of the iron) orbitals. The occupancy of the σ^*_{Fe-S} orbital (see **Figure 6**) causes remarkable differences between the doublet and quartet states. In the quartet state this orbital is singly occupied increasing the spin population on the SH⁻ ligand as well. In contrast, this orbital is unoccupied in the doublet state, which increases the electron acceptor character of the ferric ion, which allows for stronger ligand binding. This is the main reason behind the strikingly different Fe-O bond lengths obtained for among others the above mentioned quartet (3.430 Å) and doublet (2.213 Å) C_{QM} systems.

When one compares the occupation of orbitals in different spin states of a complexes, it is value to see how such differences influence the spin-state splitting of the compounds. This question is especially interesting, because the different DFT functionals have been shown to give strikingly different spin-state splittings.[89,90] In our case (see **Table 1**) it is found that the quartet and doublet states of A_{QM} and B_{QM} are predicted to be very close in energy by all functionals. However, the spin state splittings of systems C_{QM}, D_{QM} and E_{QM} are predicted differently by the various functionals (with differences up to 10 kcal/mol). These are the complexes in which in the quartet state an electron is promoted to the σ^*_{Fe-S} orbital. This finding is in line with a recent work, which implied that largest differences between hybrid and pure functionals in predicting spin-state splittings can be expected for systems in which an electron is promoted from a non-bonding orbital to an antibonding one.[91]

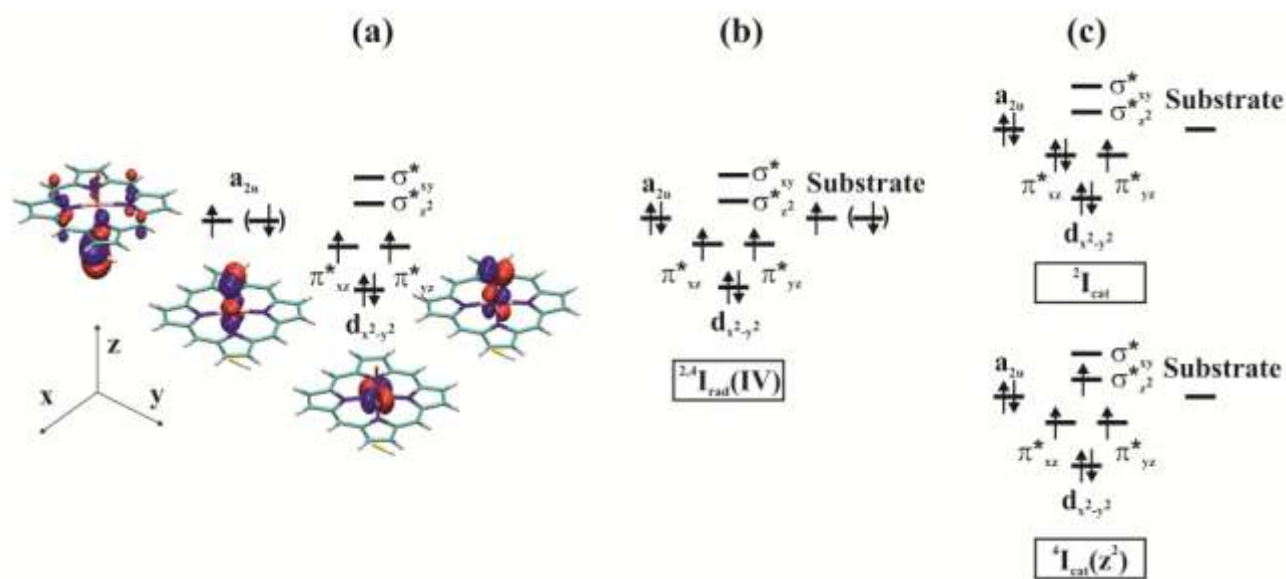**Figure 5.** Occupancy of the highest lying molecular orbitals of (a) Compound I (b) radical intermediate (B_{QM}) suggested by gas-phase calculations; (c) cationic intermediates ($B_{QM/MM}$) suggested by QM/MM calculations (see below). In the case of Compound I, the shape of the molecular orbitals is also shown. Color coding: pink: iron; dark blue: nitrogen; light blue: carbon, yellow: sulfur; white: hydrogen.
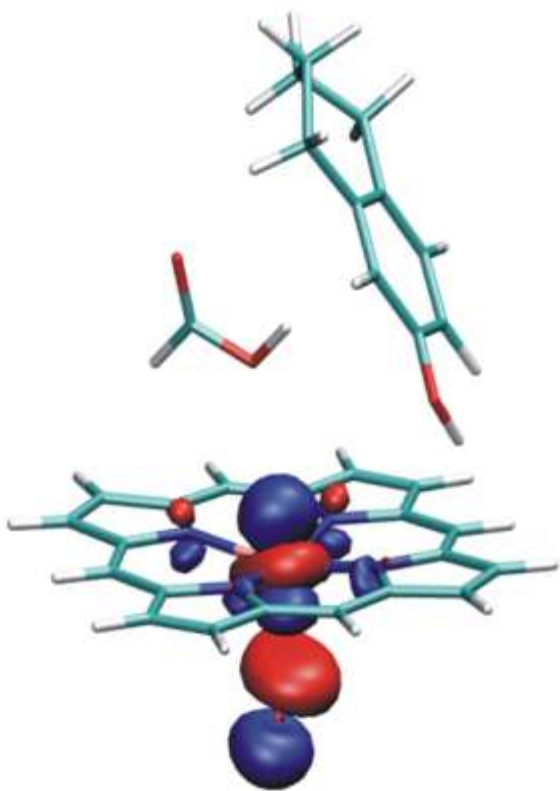
Figure 6. Singly occupied σ^*_{Fe-S} orbital in the $^4D_{QM}$ system (d_{z2} orbital of the ferric ion).

In summary the QM-only calculations suggest that the deformylation process is an energetically favorable process, which can occur via an $^{2,4}I_{rad}(IV)$ type intermediate involving low energy barriers (except for the C_{QM}=>D_{QM} conversion). Therefore we investigated this pathway using a hybrid QM/MM technique in order to take the protein environment into account.

3.2 QM/MM calculations.

3.2.1. The energy landscape. Correct sampling of enzyme configurations is essential to obtain reliable results when modelling enzymes.[92,93] However, due to the large number of energy and

gradient calculations, these methods (e.g. QM/MM umbrella sampling[93,94]) are mostly used with relatively low QM levels, such as semi-empirical methods, which cannot be used to reliably model cytochrome P450 enzymes. It should be noted, that CPMD method by Car and Parrinello[95], and the newly emerging GPU-accelarated ab initio MD simulations[96,97] will hopefully bring a breakthrough to the QM/MM studies of CYPs. However, in the present study, in order to approximate to a sufficient amount of sampling, QM/MM potential energy profiles have been calculated using three randomly chosen structures derived from the last nanosecond of the 2 ns long MD trajectory similarly to other studies.[86,98,99] These starting structures were snapshots of the enzyme, thus differed in many minor details which could arise due to thermal motion of the atoms. A very important difference from the QM/MM study published earlier[26] was that we observed no hydrogen bond between the side chain of Thr_{310} and the formyl group of the substrate during the entire length of the MD simulation, since the catalytic water molecule remained between the Thr_{310} side chain and the Fe-O moiety forming hydrogen bonds to both of them. Using adiabatic mapping we constrained the system to go along the investigated pathways shown in **Scheme 4**, and the stationary points of the profiles were investigated similarly to the QM-only model systems. Identical frames of the MD trajectory were used for the doublet and quartet states, thus three parallel doublet and quartet profiles were calculated, each pair (e.g. profile 1 in the doublet and quartet states) starting from the same frame of the MD trajectory.

The energy gain of the whole aromatization process is around 100 kcal/mol (**Figure 7**) and is much higher than the value calculated for the hydroxylation reaction in $P450_{cam}$ (around 40 kcal/mol),[86] which is due to the release of energy associated to the aromatization of ring A, as was argued above. The doublet and quartet systems behave slightly differently as the hydroxo complex (**$B_{QM/MM}$**) was found to be a local minimum on the doublet PES, while it did not exist in

the quartet state. This difference between the two spin states might be explained by their different electronic structure, as will be discussed later.

From the data presented in **Figure 7** it can be seen that the results are not sensitive to the size of the basis set. The rate-determining step of the reaction is 1β-hydrogen abstraction, whose barrier of is about 3.5-4.0 kcal/mol in the doublet state, which is in the same range as value obtained in the QM-only calculation and it is about 2 kcal/mol higher in the quartet state. At the end of the 1β-hydrogen abstraction step, we carried out an unconstrained geometry optimization of the structures. This led to the **B$_{QM/MM}$** structure in all three doublet profiles, which can be converted via low energy barriers (about 2-3 kcal/mol) to the **C$_{QM/MM}$**, **D$_{QM/MM}$** or **E$_{QM/MM}$** structures (**Figure 8**) using the corresponding reaction coordinates (**Figure 1**). Since the energy barriers of the three pathways (**B$_{QM/MM}$**=>**C$_{QM/MM}$**, **B$_{QM/MM}$**=>**D$_{QM/MM}$** and **B$_{QM/MM}$**=>**E$_{QM/MM}$**) are comparable to the error of the method, the calculations suggests all of them to be very fast processes. Therefore, the **B$_{QM/MM}$** state could be regarded as a crossing point of these three pathways, but as experimentally only **D$_{QM/MM}$** has been observed so far, the route leading to **D$_{QMMM}$** should dominates the enzyme's reactivity.

In the quartet state the unconstrained optimizations led directly to the **D$_{QM/MM}$** state in the case of profiles 1 and 3 or to the **C$_{QM/MM}$** structure (profile 2) depending on the exact orientation of heme-bound hydroxyl group. The most likely reason is that the OH ligand is bound weaker to the heme in the quartet state, and is therefore able to move somewhat closer to C_1 or C_{19} atoms of the substrate in the course of the reaction coordinate calculation and form the **C$_{QM/MM}$** or the **D$_{QM/MM}$** structures directly instead of the **B$_{QM/MM}$** intermediate. Although based on the gas-phase calculations, the formation of carbon monoxide (**E$_{QM/MM}$** state) should be a fast process in the quartet state as well the unrestrained geometry optimization of the final structure of the 1β-

hydrogen abstraction process never led to this product, because C_{19}-H bond of the formyl group was not oriented towards the oxygen coordinated to the iron. We expect that in a real system where thermal motion could lead to fluctuations in the conformation of the formyl group, CO might be produced in the enzyme in the quartet as well.

At this point we would like to note that the relatively large differences between the relative energies of the corresponding states obtained from various profiles most likely originate from the shortcomings of the widely used[100] QM/MM adiabatic mapping methodology. Using adiabatic mapping we carried out a large number of QM/MM energy minimizations along the given reaction coordinates, however, in all cases the calculations always converged to the nearest local minimum on the potential energy surface. This phenomenon has been previously investigated by Warshel et al[101] and was suggested to be especially problematic in the case of a rugged potential energy surface with many local minima. The fact that the calculated barriers differ only by less than 3 kcal/mol suggests their reliability.
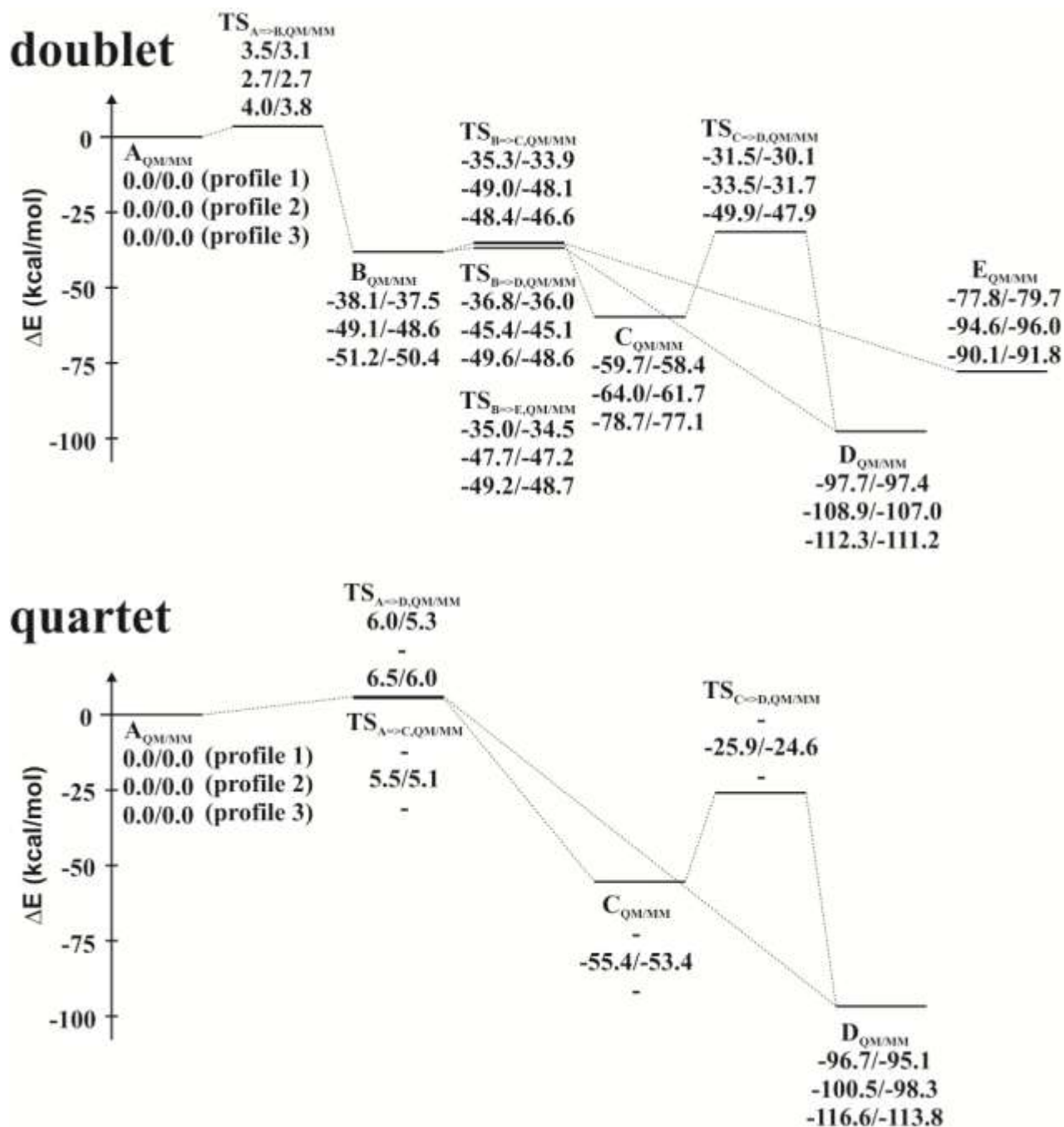
Figure 7. Energy diagram of the investigated QM/MM model systems in the doublet and quartet states. Energies are given for three profiles at the B3LYP/B1/MM (left) and B3LYP/6-311+G*/MM (right) level including dispersion correction. The representative structures of the QM region in the stationary points are taken from profile 1, selected geometrical parameters are given in **Table S4** in the SI.

$\mathbf{A_{QM/MM}}$ $\qquad$ $\mathbf{TS_{A=>B,QM/MM}}$ $\qquad$ $\mathbf{B_{QM/MM}}$

$\mathbf{TS_{B=>C,QM/MM}}$ $\qquad$ $\mathbf{TS_{B=>D,QM/MM}}$ $\qquad$ $\mathbf{TS_{B=>E,QM/MM}}$

$\mathbf{C_{QM/MM}}$ $\qquad$ $\mathbf{D_{QM/MM}}$ $\qquad$ $\mathbf{E_{QM/MM}}$

$\mathbf{TS_{C=>D,QM/MM}}$

32

Figure 8. QM region of the QM/MM model systems. The representative structures are taken from profile 1 in the doublet state and selected geometrical parameters are given in **Table S4** in the SI.

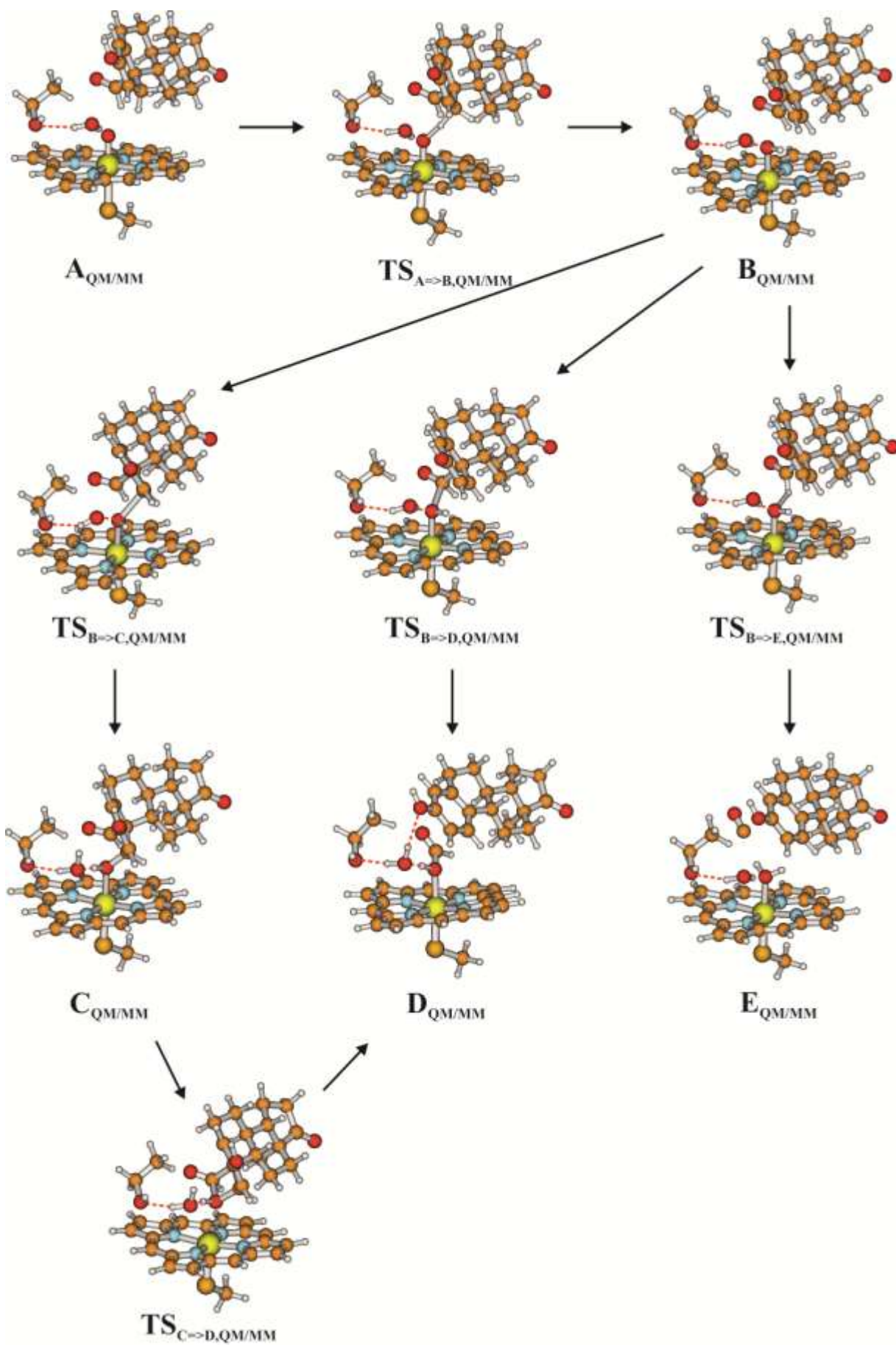At this point, it is worth examining the results of the QM/MM calculations in the light of the experimental data available on the enzyme mechanism. The known product state of aromatase is $D_{QM/MM}$: the aromatized steroid (in the case of ASD estrone) and formic acid. Neither $C_{QM/MM}$ (the 1β-hydroxylated by-product) nor $E_{QM/MM}$ (formation of CO instead of formic acid besides the aromatized estrone) have been reported in the case of the natural substrates of the enzyme. Using the techniques employed in this work it is impossible to distinguish between these pathways, as the differences in the activation energies of the various pathways is in the same range as the differences obtained for the snapshots studied. Out of the three pathways presented herein most likely the pathway leading to $D_{QM/MM}$ dominates. The formation of the 1β-hydroxylated by-product of 19-nor-ASD is easily explained by our calculations, and we believe that small amounts might be produced in the case of ASD. $C_{QM/MM}$, either undergoes intramolecular rearrangement and C_{10}-C_{19} bond scission (although the calculated barrier is about 30 kcal/mol, thus this process is not very likely) or most likely a very facile non-enzymatic pathway exists, similarly to the case of the 2β-hydroxylated derivative[13] as suggested in **Scheme 5.**

We examined the effect of the DFT functional on the obtained results using single-point calculations on a selected doublet and quartet QM/MM profile (profile 1, see **Table 2**). The conclusions are basically the same as in the case of QM-only calculations: the trends obtained by B3LYP are more or less reproduced by the other functionals, although the numerical value of the

relative energies depends extensively on the DFT functional. The M06 functional gave rather similar relative energies to B3LYP, and the OLYP values were the most different. 1β-hydrogen abstraction is suggested to be a very fast conversion by the B3LYP, M06 and ωB97X-D, while TPSSh and OLYP gave slightly higher activation energies. The energy barrier for the **C**$_{QM/MM}$=>**D**$_{QM/MM}$ conversion was found to be similar to the unimolecular process studied in the gas-phase in the absence of the heme ring. This implies that the enzymatic environment does not catalyze this process.

Table 2. Effect of the applied DFT functional on the relative energies (in kcal/mol) of the different systems in QM/MM calculations[a]

	model	B3LYP[b]	M06	TPSSh	OLYP	ωB97X-D
	A$_{QMMM}$	0.0	0.0	0.0	0.0	0.0
	TS$_{A=>B,QMMM}$	3.1	2.9	7.5	13.2	3.1
	B$_{QMMM}$	-37.5	-37.7	-25.4	-9.3	-45.1
	TS$_{B=>C,QMMM}$	-33.9	-36.2	-20.9	-2.9	-41.1
profile 1	**C**$_{QMMM}$	-58.4	-59.8	-39.6	-15.4	-69.8
(doublet)	**TS**$_{C=>D,QMMM}$	-30.1	-32.2	-14.9	6.0	-40.7
	TS$_{B=>D,QMMM}$	-36.0	-36.3	-23.6	-7.1	-40.7
	D$_{QMMM}$	-97.4	-101.5	-89.0	-79.6	-108.8
	TS$_{B=>E,QMMM}$	-34.5	-34.3	-24.4	-10.7	-41.8
	E$_{QMMM}$	-79.7	-84.8	-68.0	-58.6	-91.3
profile 1	**A**$_{QMMM}$	0.0	0.0	0.0	0.0	0.0
(quartet)	**TS**$_{A=>D,QMMM}$	5.3	4.8	8.1	12.1	4.7
	D$_{QMMM}$	-95.1	-107.8	-82.6	-78.6	-107.7

[a]Energies are derived from single-point calculations using the 6-311+G* basis set

[b]B3LYP energies are dispersion corrected.

In summary, the QM/MM energy profiles suggest that that aromatase can efficiently catalyze the deformylation of the enolized 19-oxo-ASD ligand in a pathway involving Compound I. The energy barriers are very low in both spin states except for the **C**$_{QM/MM}$=>**D**$_{QM/MM}$ conversion

which is a highly exothermic process, and single point calculations using other DFT functionals support this hypothesis.

3.2.2. Electronic structure of the obtained stationary points.

The most striking difference between the QM-only and the hybrid QM/MM calculations is that the energy barrier for 1β-hydrogen abstraction is almost the same in the doublet QM-only and QM/MM, and quartet QM-only systems, but it is higher in the quartet QM/MM systems. Although, this difference is not substantial, it might originate from a slightly different electronic structure. To clarify this issue we carefully analyzed the obtained natural charges and spin populations (see **Figure 9** and **Table S5** in the SI).
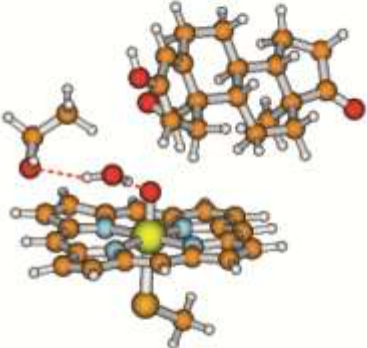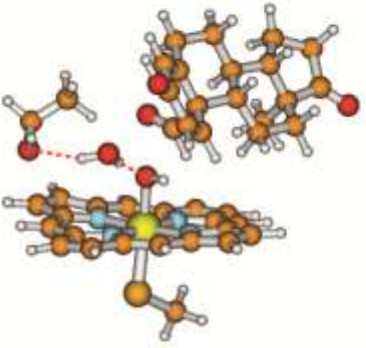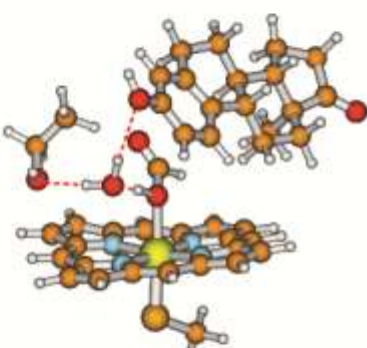
	$A_{QM/MM}$		$B_{QM/MM}$		$D_{QM/MM}$	
	Q	SP	Q	SP	Q	SP
Fe	0.98/0.97	1.24/1.19	0.94/–	0.93/–	0.88/1.20	0.94/2.41
	1.00/0.99	1.31/1.30	0.93/–	0.88/–	0.89/1.23	0.94/2.44
	0.98/0.98	1.30/1.26	0.94/–	0.94/–	0.88/1.20	0.94/2.41
O	-0.63/-0.62	0.77/0.80	-0.99/–	0.01/–	-0.73/-0.75	0.01/0.04
	-0.68/-0.68	0.68/0.71	-0.93/–	0.09/–	-0.74/-0.75	0.01/0.04
	-0.68/-0.68	0.69/0.74	-0.99/–	0.01/–	-0.73/-0.75	0.01/0.04

	$A_{QM/MM}$		$B_{QM/MM}$		$D_{QM/MM}$	
Porphyrin	-0.45/-0.46	-0.29/0.22	-0.81/–	-0.01/–	-0.76/-0.74	0.00/0.14
	-0.65/-0.66	-0.04/-0.03	-0.84/–	-0.02/–	-0.75/-0.73	0.00/0.14
	-0.55/-0.54	-0.16/0.11	-0.80/–	-0.01/–	-0.75/-0.75	0.00/0.14
MeS⁻	-0.48/-0.46	-0.14/0.16	-0.41/–	0.05/–	-0.25/-0.51	0.05/0.40
	-0.57/-0.57	0.01/0.01	-0.49/–	0.05/–	-0.29/-0.56	0.04/0.37
	-0.48/-0.49	-0.09/0.09	-0.43/–	0.05/–	-0.26/-0.52	0.05/0.40
Substrate	0.58/0.59	-0.58/0.61	0.85/–	0.00/–	0.01/0.00	0.00/0.00
	0.87/0.85	-0.90/0.92	0.83/–	0.00/–	0.02/0.02	0.00/0.00
	0.73/0.73	-0.75/0.78	0.85/–	0.00/–	0.01/0.02	0.00/0.00

Figure 9. Natural charges (Q) in e and spin populations (SP) for selected atoms and fragments in QM/MM calculations at the B3LYP/6-311+G*/MM level. Data are given for all three profiles (shown in different rows) in the doublet (left) and quartet (right) states.

The most conspicuous differences in the spin population and charge distribution values derived from QM-only and QM/MM calculations are observed in the case of the **A** and **B** states. As long as the A_{QM} state contains obviously a complex of Compound I with zero spin population on the uncharged substrate molecule, there is a strong charge transfer interaction between the substrate and the iron complex in the $A_{QM/MM}$ models indicated by the considerable positive charge and spin population on the substrate molecule. The charge transfer interaction is very sensitive for the environment demonstrated by the different charge and spin population values in the three-three parallel doublet and quartet profiles. The spin population on the porphyrin ring and the MeS$^-$ ligand is much lower than in the A_{QM} system, which suggests that the electron transfer is directed from the π-system of the substrate to the a_{2u} porphyrin orbital mixing with the p_z orbital of the sulfur. This is not surprising taken the highly electrophilic character of Compound I and the very electron rich π-system of the dienol ligand. This interaction is visualized in **Figure 10**. In the case of the doublet and quartet profiles with lowest barriers (profile 2) this interaction is so strong, that the $A_{QM/MM}$ system contains practically a Compound II complex, which is the one electron reduced form of Compound I, with a doubly occupied a_{2u} porphyrin orbital. This result is in accordance with another QM/MM study, in which Compound II was incorporated in the mechanism of 1β-hydrogen abstraction instead of Compound I.[26] In the other profiles the nature of the $A_{QM/MM}$ state is best described as a mixture of Compound I and Compound II characters with the latter one being stronger. Similar partial delocalization of the electrons of Compound I to nearby residues has been reported earlier for peroxidases and cytochrome P450 enzymes.[102]
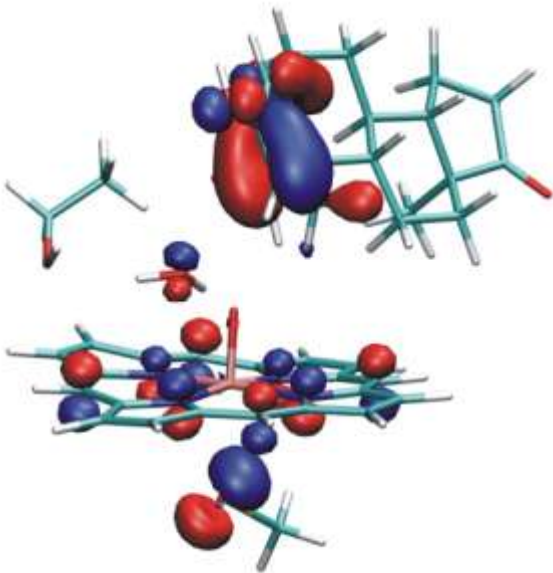
Figure 10. Singly occupied beta Kohn-Sham orbital derived from the $A_{QM/MM}$ state of profile 1 in the double state showing the interaction between the π orbital of the substrate molecule and the a_{2u} porphyrin orbital.

In order to elucidate the origin of these differences in the electronic structure, to determine whether these are real effects of the different geometry and the polarizing environment or artificial influences of self-interaction error, we have addressed the questions from two sides. First, we carefully analyzed the results obtained by various functionals in order to identify the likeliness of significant self-interaction error. Second, we have investigated the effect of the protein environment and of the catalytic water molecule on the spin distribution of the systems (see the discussion in Chapter 3.2.3).

The natural charges and spin populations calculated with other DFT functionals are summarized in **Table S6** in the SI for the $A_{QM/MM}$ and $B_{QM/MM}$ structures of profile 1. Basically all of these calculations suggest the same cationic configuration for $B_{QM/MM}$ with zero spin

population and about 0.85 natural charge on the substrate. For the $\mathbf{A_{QM/MM}}$ system the M06, TPSSh and OLYP functionals give the same spin distribution as B3LYP with about 0.6 e natural charge and ±0.6 spin population on the substrate depending on the multiplicity. In contrast, the ωB97X-D functional suggests a clear Compound II character for the $\mathbf{A_{QM/MM}}$ system with about 0.96 e natural charge and ±1.0 spin population on the substrate. Since the ωB97X-D functional is a long-range corrected functional with -1/r decay of the exchange-correlation potential, it is free of long-range self-interaction.[69] Therefore, our findings might imply that the mixed character of the reactant state is due to self-interaction error of the DFT functionals with no long-range correction. However, this statement has to be taken carefully, as the reliability and accuracy of the ωB97X-D functional has not been tested yet on heme systems, unlike B3LYP which has been thoroughly tested and is known to give reliable results for cytochrome P450 enzymes.[60]

Based on these electronic features we can explain the differences in the barrier heights for the 1β-hydrogen abstraction in the QM-only and QM/MM calculations. In the QM-only calculations the barriers in the doublet and quartet states are almost the same as identical changes occur in the two states: during hydrogen atom transfer an electron is placed from the π-orbital of the substrate to the a_{2u} porphyrin orbital. In contrast, according to the QM/MM calculations first electron transfer and then hydrogen atom transfer occurs. In the $\mathbf{^2B_{QM/MM}}$ state there is only one unpaired electron, which is located on one of the π^*_{xz} or π^*_{yz} orbitals ($\mathbf{^2I_{cat}}$ in **Figure 5**), but in the quartet state one of the three unpaired electrons has to be placed on the higher-lying σ^*_{Fe-S} orbital in the course of the 1β-hydrogen abstraction ($\mathbf{^4I_{cat}(z^2)}$ in **Figure 5**), which increases the barrier compared to the doublet state. Since this orbital is unoccupied in the doublet state the hydroxyl ion binding can be stronger than in the quartet system and this is why we were able to obtain the $\mathbf{B_{QM/MM}}$ state as a local minimum only in the doublet state.

The charge and spin distributions are similar in the $C_{QM/MM}$, $D_{QM/MM}$ and $E_{QM/MM}$ systems, thus for the sake of simplicity, only those of the latter are depicted in **Figure 9**. The electronic configuration in these systems is the same as in the C_{QM}, D_{QM} and E_{QM} models. There are five electrons in the d-block of the iron center, from which one is unpaired in the doublet state and located on one of the d_{xz} and d_{yz} orbitals, and three electrons are unpaired in the quartet state located on the d_{xz}, d_{yz} and the σ^*_{Fe-S} (d_{z2} of the iron ion) orbitals. This difference causes significant spin population on the MeS$^-$ ligand in the quartet $C_{QM/MM}$, $D_{QM/MM}$ and $E_{QM/MM}$ structures, which are also characterized by longer Fe-O distances than the doublet structures. (2.4 Å vs. 2.1 Å; see **Table S4**). However, both of these values are much shorter than the values found in QM-only structures, which most likely originates from the fact that the QM/MM calculations were carried out using dispersion correction, and the free movement of the ligands is hindered by the protein environment in contrast to the QM-only calculations.

3.2.3. Estimation of the overall catalytic effect of human aromatase.

The results presented above show that Compound I can be the active agent that is responsible for the cleavage of the C_{10}-C_{19} bond and for the aromatization reaction in accordance with experimental studies.[21,22] Now, we would like to get some idea of the catalytic effect of aromatase, and compare it to relevant data obtained for other enzymes earlier. As described in detail by Warshel et al.[103] numerous proposal have been put forward in order to explain the origin of enzymatic catalysis, and reliable quantification of this effect requires rigorous, energy based treatment and a proper definition of the reference state to which the catalytic effect of the enzyme could be compared to.

When one tries to understand the catalytic effect of aromatase, it should be considered what exactly the role of aromatase is in the reaction. It seems highly unlikely that if the ingredients depicted in **Scheme 1** were put together in solution, the reaction aromatization reaction would occur. Thus aromatase has an essential role in activation of molecular oxygen by breaking the O=O bond and producing the highly reactive Compound I species, which can carry out the three oxidation reactions in the three catalytic cycles. However, at this point the question arises, is the enzyme's role solely the formation of Compound I or does it electrostatically catalyze the reaction compared to solution? I.e. is the barrier of the reaction lower in the enzyme than it would in solution?

In order to answer this question, it is valuable to consider the arguments put forward in ref. 103 and to define a proper reference reaction to which the enzyme's reactivity could be compared to. In the present case it is impossible to define a proper solvent catalyzed reference reaction for the aromatization of androstenedione. However, by defining a hypothetical reference reaction, one could determine the barrier of the reaction in solution using the enzyme mechanism (as was suggested in ref. 103) and compare the activation energy in the enzyme and in solution using the same mechanism. This "mechanism-filtered" reference reaction in water is only a hypothetical one and it is simply used to estimate the electrostatic stabilization of the transition state by protein pre-organized dipole in contrast to random water dipoles.

Therefore, we have carried out a number of QM/MM single point calculations on the reactant and intermediate states of the reaction ($A_{QM/MM}$ and $B_{QM/MM}$) and on the transition state of the 1β-hydrogen abstraction step. ($^2TS_{A \Rightarrow B,QMMM}$) in the doublet state (**Table 3**). Besides the overall catalytic effect of the enzyme, we also wanted to address the specific contribution of the catalytic water molecule and of the Asp$_{309}$ side-chain to the reaction. Therefore the single point

calculations were carried out in the absence and presence of the enzyme (as represented by point charges) and the catalytic water molecule (which was included in the QM region when it was present). The activation energy of the reaction in solution was approximated by continuum solvent calculations.

In the absence of point charges the relative energy of the transition state increases from 6.6 kcal/mol to 12.6 kcal/mol in vacuo. When a PCM model is used instead of the electrostatic environment of the enzyme (point charges) the barrier drops to 8 kcal/mol. The larger effect of the PCM model on the relative energy of the transition state compared to the calculation without any point charges is due to the ionic character of the TS: which is considerably stabilized by the aqueous environment, and even more by the protein environment. This can be understood, as the protein environment could efficiently help in the stabilization of the dipole moment rising inside the QM region in the course of the formation of a cationic intermediate.[104] From this it can be concluded that the enzyme slightly (by about 1.5 kcal/mol) catalyzes the reaction compared to solution, which is much smaller than values reported for other enzymes (5.7-34 kcal/mol).[103] The small effect of the enzyme on reaction barrier compared to solution also suggests that the enzyme's primary role is the formation of Compound I by activation of O_2, and no the electrostatic stabilization of the transition state. Of course the number found in our study have to be treated cautiously, as (1) these numbers are within the accuracy of B3LYP calculation, (2) continuum solvent methods have their own limitations (e.g. the effect of cavity shape and size on the solvation free energy, the real dielectric constant of the system is not known), but they do suggest that the role of the enzyme is not simply the generation of Compound I, but slightly decreasing the activation energy by electrostatic interactions.

Removal of the catalytic water molecule from the system does not change the energy barrier or the energy gain of the reaction considerably. The idea that the deprotonated Asp_{309} residue located in the vicinity of the ligand could have a significant role in the stabilization of the ionic transition state is not supported either. Single point calculations done with the artificially neutralized Asp_{309} residue show only a minor effect (**Table 3**) compared to the overall effect of the electrostatic environment of the protein.

The results of single point calculations summarized in **Table 3** can also help a better understanding of origin of the differences in the electronic structure of the studied species. The most important finding is that the neglect of the catalytic water molecule and removal of the point charges from the calculation do not change the cationic character of the $^2\mathbf{B_{QM/MM}}$ intermediate state. In contrast, the charge transfer interaction in the $^2\mathbf{A_{QM/MM}}$ state is sensitive for these factors. In the absence of the catalytic water the natural charge and spin population on the substrate molecule considerably decrease, and it becomes even more remarkable in the absence of the point charges. In the absence of both of them the $^2\mathbf{A_{QM/MM}}$ state has a clear Compound I character resembling the $^2\mathbf{A_{QM}}$ state. Removal of the point charges and catalytic water molecule from the $^2\mathbf{B_{QM/MM}}$ state leads to a slight decrease of the charge on the substrate molecule, but it does not alter its cationic character with about zero spin population. These results imply that it is the enzyme's active site architecture that is primarily responsible for the cationic character ($^2\mathbf{I_{cat}}$ in **Figure 5**) of the intermediate, which is further stabilized by the electrostatic interactions with the protein environment and the catalytic water molecule. This is in sharp contrast to the $^2\mathbf{B_{QM}}$ model where free movement of the fragments allowed the formation of a more favorable intermediate with a radical character. These findings are in line with previous works, which

showed that spatial reasons may account for the formation of cationic intermediates, which are usually characterized by higher barriers than the pathways involving radical intermediates.[29,30]

Table 3. Natural charges (Q, in e), spin populations (SP) and relative energies (in kcal/mol) in the absence of the catalytic water molecule and/or MM point charges and the role of the protonation state of Asp_{309} residue at the B3LYP/6-311+G* level.[a]

state	catalytic water MM point charges charge on Asp_{309} atom or group	present present -1 Q	 SP	present absent - Q	 SP	absent present -1 Q	 SP	absent absent - Q	 SP	present present 0^b Q	 SP	present no => PCM(water) - Q	 SP
$^2A_{QMMM}$	Fe	0.98	1.24	0.96	1.20	0.94	1.12	0.93	1.09	0.97	1.24	0.97	1.25
	O	-0.63	0.77	-0.58	0.81	-0.52	0.90	-0.48	0.93	-0.62	0.78	-0.61	0.76
	MeS⁻	-0.48	-0.14	0.00	-0.67	-0.45	-0.22	0.01	-0.72	-0.41	-0.22	-0.11	-0.54
	porphyrin	-0.45	-0.29	-0.45	-0.27	-0.38	-0.42	-0.46	-0.31	-0.30	-0.43	-0.22	-0.49
	substrate	0.58	-0.58	0.08	-0.08	0.39	-0.39	0.00	0.01	0.36	-0.36	0.00	0.01
	E_{rel} (QM region)	0.0		0.0		0.0		0.0		0.0		0.0	
$^2TS_{A=>B,QMMM}$	Fe	1.02	1.44	1.01	1.59	1.01	1.41	1.00	1.55	1.03	1.50	1.02	1.50
	O	-0.72	0.43	-0.77	0.32	-0.64	0.45	-0.68	0.35	-0.74	0.37	-0.77	0.39
	MeS⁻	-0.44	-0.04	-0.20	-0.17	-0.45	-0.05	-0.20	-0.21	-0.40	-0.07	-0.36	-0.07
	porphyrin	-0.59	-0.04	-0.59	-0.08	-0.60	-0.05	-0.58	-0.12	-0.56	-0.05	-0.60	-0.04
	substrate	0.74	-0.80	0.56	-0.67	0.68	-0.76	0.46	-0.57	0.68	-0.77	0.71	-0.78
	E_{rel} (QM region)	6.6		12.6		6.4		15.3		7.7		8.0	
$^2B_{QMMM}$	Fe	0.94	0.93	0.88	0.88	0.94	0.94	0.89	0.90	0.93	0.93	0.92	0.92
	O	-0.99	0.01	-1.00	0.01	-0.95	0.02	-0.95	0.03	-1.00	0.01	-1.01	0.02
	MeS⁻	-0.41	0.05	-0.21	0.14	-0.43	0.06	-0.22	0.14	-0.39	0.06	-0.37	0.07
	porphyrin	-0.81	-0.01	-0.85	-0.04	-0.82	-0.01	-0.84	-0.04	-0.80	-0.01	-0.83	-0.02
	substrate	0.85	0.00	0.77	0.00	0.82	0.00	0.69	-0.04	0.84	0.00	0.87	0.00
	E_{rel} (QM region)	-29.6		-20.1		-25.4		-12.2		-28.3		-31.1	

[a]The structures are taken from the doublet profile 1.

[b]The charges on the side chain atoms of Asp_{309} have been modified to account for its neutral character in accordance with the patch residue used in CHARMM for the buffer region. (C_β: -0.28 => 0.0; O: -0.76 => -0.40)

4. CONCLUSION

In this work we have studied the last step of the aromatization of androgens to estrogens catalyzed by aromatase using QM-only and QM/MM calculations. We propose that in the final catalytic cycle Compound I (or Compound II) abstracts the 1β–hydrogen atom from the enol form 19-oxo-ASD (which is produced by 2β–hydrogen abstraction from the ligand by the ferrous superoxo (ferric peroxo) complex[46]), which leads to the scission of the C_{10}-C_{19} bond, and the formation of estrone and formic acid. The rate-determining step of the reaction is 1β-hydrogen abstraction, whose barrier is in the range of 3-6 kcal/mol, which is much lower than the activation energy of many CYP-catalyzed hydrogen abstraction reactions, therefore the reaction is supposed to be very fast. According to our results the human aromatase may catalyze the reaction via Compound I in three pathways all leading to estrone. Most likely the direct pathway dominates the enzyme's reactivity, but besides formic acid the production of a small amount of carbon monoxide cannot be excluded. Our calculations provide a plausible explanation for the most recent resonance Raman and KSIE measurements which suggested the involvement if Compound I in the third catalytic sub-cycle of human aromatase.[27,28]

A significant difference has been found in the reaction mechanism as predicted by QM-only and QM/MM calculations. According to the QM-only calculations 1β-hydrogen is abstracted by Compound I and it leads to radical intermediates. In QM/MM calculations all of the used DFT functionals suggest that the reaction involves cationic intermediates, but the oxidant has a mixed Compound I and Compound II character according to DFT functionals with no long-range correction, and clearly Compound II character according to the dispersion and long-range corrected ωB97X-D functional. This may suggest that the real configuration of the oxidant is clearly Compound II and the mixed character is derived from self-interaction error. The

Compound II character of the oxidant arises as a consequence of the polarizing effect of the protein environment and of the catalytic water molecule generated by the formation of Compound I from Compound 0.

We have estimated the catalytic power of the enzyme using a continuum solvent model, by comparing the activation energy of the enzyme catalyzed reaction to a hypothetical reference reaction in aqueous solution at the geometry of the enzyme-catalyzed reaction. It was found that the electrostatic environment of the enzyme only slightly (by about 1.5 kcal/mol) decreases the energy of activation of the reaction compared to the reaction in solvent, thus most likely the most important role of the enzyme is the generation of Compound I and the orientation of the substrate in the active site.

All the used DFT functionals predict the same trends (both for energetics and changes in the electronic structure) in QM-only and QM/MM calculation. However, the numerical values of the obtained relative energies differed considerably (as much as up to 30 kcal/mol for the overall reaction energy), which implies that when direct comparison with experimental data is possible, a very careful analysis of the reliability of the DFT functional is needed.

SUPPORTING INFORMATION. Selected geometrical parameters and fragment charges for QM-only optimized structures using various basis sets, and details of all reactant, TS and product complexes obtained in QM/MM calculations. Furthermore, total energies and Cartesian coordinates of all QM-only structures, geometries of the QM region of the QM/MM structures belonging to profile 1, and MM parameters for the 19-oxo-ASD are also provided in the SI together with tables and figures indicated in the text.

REFERENCES

1   Groves, J. T. Models and Mechanisms of Cytochrome P450 Action In *Cytochrome P450*, third edition; Ortiz de Montellano, P. R., Ed.; Kluwer Academic/Plenum Publishers: New York, 2005; pp 1-43.

2   Sligar, S. G. Glimpsing the Critical Intermediate in Cytochrome P450 Oxidations. *Science* **2010**, *330*, 924-925.

3   Rittle, J.; Green, M. T. Cytochrome P450 Compound I: Capture, Characterization, and C-H Bond Activation Kinetics. *Science* **2010**, *330*, 933-937.

4   Brodie, A. M. H.; Njar, V. C. O. Aromatase Inhibitors and their Application in Breast Cancer Treatment. *Steroids* **2000**, *65*, 171-179.

5   Baum, M.; Budzar, A. U.; Cuzik, J.; Forbes, J.; Houghton, J. H.; Klijn, J. G.; Sahmoud, T. Anastrozole Alone or in Combination With Tamoxifen Versus Tamoxifen Alone for Adjuvant Treatment of Postmenopausal Women with Early Breast Cancer: First Results of the ATAC Randomised Trial. *Lancet* **2002**, *359*, 2131-2139.

6   Bonneterre, J.; Buzdar, A.; Nabholtz, J. M.; Robertson J. F.; Thurlimann, B.; von Euler, M.; Sahmoud, T.; Webster, A.; Steinberg, M. Anastrozole Is Superior to Tamoxifen as First-Line Therapy in Hormone Receptor Positive Advanced Breast Carcinoma. *Cancer* **2001**, *92*, 2247-2258.

[7]   Brueggemeier R. W. Aromatase Inhibitors in Breast Cancer Therapy. *Expert Rev. Anticancer Ther.* **2002**, *2*, 181-191.

[8]   Ryan, K. J. Conversion of Androstenedione to Estrone by Placental Microsomes. *Biochim. Biophys. Acta* **1958**, *27*, 658-659.

[9]   Ortiz de Montellano, P. R.; De Voss, J. J. Substrate Oxidation by Cytochrome P450 Enzymes. In *Cytochrome P450*, third edition; Ortiz de Montellano, P. R., Ed.; Kluwer Academic/Plenum Publishers: New York, 2005; pp 183-245.

[10]   Groves, J. T.; McClusky, G. A. Aliphatic Hydroxylation via Oxygen Rebound. Oxygen Transfer Catalyzed by Iron. *J. Am. Chem. Soc.* **1976**, *98*, 859-861.

[11]   Osawa, Y.; Shibata, K.; Rohrer, D.; Weeks, C.; Duax, W. L. Reassignment of the Absolute Configuration of 19-Substituted 19-Hydroxysteroids and Stereomechanism of Estrogen Biosynthesis. *J. Am.Chem. Soc*. **1975**, *97*, 4400-4402.

[12]   Arigoni, D.; Battaglia, R.; Akhtar, M.; Smith, T. Stereospecificity of Oxidation at C-19 in Oestrogen Biosynthesis. *J. Chem. Soc., Chem. Commun*. **1975**, No. 6, 185-186.

[13]   Hosoda, H.; Fishman, J. Unusually Facile Aromatization of 2.Beta.-Hydroxy-19-Oxo-4-Androstene-3,17-Dione to Estrone: Implications in Estrogen Biosynthesis. *J. Am. Chem. Soc*. **1974**, *96*, 7325-7329.

[14]   Goto, J.; Fishman, J. Participation of a Nonenzymatic Transformation in the Biosynthesis of Estrogens from Androgens. *Science* **1977**, *195*, 80-81.

[15]   Hahn, E. F.; Fishman, J. Immunological Probe of Estrogen Biosynthesis: Evidence for the 2 Beta-Hydroxylative Pathway in Aromatization of Androgens. *J. Biol. Chem*. **1984**, *259*, 1689-1694.

[16]   Morand, P.; Williamson, D. G.; Layne, D. S.; Lompa-Krzymien, L.; Salvador, J. Conversion of an Androgen Epoxide into 17*β*-Estradiol by Human Placental Microsomes. *Biochemistry* **1975**, *14*, 635-638.

[17]   Mastalerz, H.; Morand, P. Acid- and Base-Catalysed Reactions of 4*β*,5*β*- and 4*α*,5*α*-Epoxyandrostane-3,17,19-Trione. *J. Chem. Soc., Perkin Trans. 1*, **1982**, No. 0, 2611-2615.

[18]   Brodie, H. J.; Kripalani, K. J.; Possanza, G. Mechanism of Estrogen Biosynthesis. VI. The Stereochemistry of Hydrogen Elimination at C-2 During Aromatization. *J. Am. Chem. Soc*. **1969**, *91*, 1241-1242.

[19]   Fishman, J.; Guzik, H.; Dixon, D. Stereochemistry of Estrogen Biosynthesis. *Biochemistry* **1969**, *8*, 4304-4309.

[20]   Akhtar, M.; Calder, M. R.; Corina, D. L.; Wright, J. N. Mechanistic Studies on C-19 Demethylation in Oestrogen Biosynthesis. *Biochem. J*. **1982**, *201*, 569-580.

[21]   Gantt, S. L.; Denisov, I. G.; Grinkova, Y. V.; Sligar, S. G. The Critical Iron-Oxygen Intermediate in Human Aromatase. *Biochem. Biophys. Res. Commun.* **2009**, *387*, 169−173.

[22]   Davydov, R.; Makris, T. M.; Kofman, V.; Werst, D. E.; Sligar, S. G.; Hoffman, B. M. Hydroxylation of Camphor by Reduced Oxy-Cytochrome P450cam: Mechanistic Implications of

EPR and ENDOR Studies of Catalytic Intermediates in Native and Mutant Enzymes. *J. Am. Chem. Soc.*, **2001**, *123*, 1403–1415.

[23] Graham-Lorence, S.; Amarneh, B.; White, R. E.; Peterson, J. A.; Simpson, E. R. A Three-Dimensional Model of Aromatase Cytochrome P450. *Protein Sci.* **1995**, *4*, 1065-1080.

[24] Tosha, T.; Kagawa, N.; Ohta, T.; Yoshioka, S.; Waterman, M. R.; Kitagawa, T. Raman Evidence for Specific Substrate-Induced Structural Changes in the Heme Pocket of Human Cytochrome P450 Aromatase during the Three Consecutive Oxygen Activation Steps. *Biochemistry* **2006**, *45*, 5631-5640.

[25] Hackett, J. C.; Brueggemeier, R. W.; Hadad, C. M. The Final Catalytic Step of Cytochrome P450 Aromatase: A Density Functional Theory Study. *J. Am. Chem. Soc.* **2005**, *127*, 5224-5237.

[26] Sen, K.; Hackett, J. C. Coupled Electron Transfer and Proton Hopping in the Final Step of CYP19-Catalyzed Androgen Aromatization. *Biochemistry* **2012**, *51*, 3039–3049.

[27] Mak, P. J.; Luthra, A.; Sligar, S. G.; Kincaid, J. R. Resonance Raman Spectroscopy of the Oxygenated Intermediates of Human CYP19A1 Implicates a Compound I Intermediate in the Final Lyase Step. *J. Am. Chem. Soc.* **2014**, *136*, 4825−4828.

[28] Khatri, Y.; Luthra, A.; Duggal, R.; Sligar, S. G. Kinetic Solvent Isotope Effect in Steady-State Turnover by CYP19A1 Suggests Involvement of Compound 1 for both Hydroxylation and Aromatization Steps. *FEBS Lett.* **2014**, *588*, 3117-3122.

29   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.* **2005**, *105*, 2279-2328.

30   Shaik, S.; Cohen, S.; Wang, Y.; Chen, H.; Kumar, D.; Thiel, W. P450 Enzymes: Their Structure, Reactivity, and Selectivity - Modeled by QM/MM Calculations *Chem. Rev.* **2010**, *110*, 949–1017.

31   Sivaramakrishnan, S.; Ouellet, H.; Matsumura, H.; Guan, S.; Moënne-Loccoz, P.; Burlingame, A. L.; Ortiz de Montellano, P. R. Proximal Ligand Electron Donation and Reactivity of the Cytochrome P450 Ferric–Peroxo Anion. *J. Am. Chem. Soc.* **2012**, *134*, 6673–6684.

32   Ouellet, H.; Guan, S.; Johnston, J. B.; Chow, E. D.; Kells, P. M.; Burlingame, A. L.; Cox, J. S.; Podust, L. M.; de Montellano, P. R. Mycobacterium Tuberculosis CYP125A1, a Steroid C27 Monooxygenase that Detoxifies Intracellularly Generated Cholest-4-en-3-one. *Mol. Microbiol.* **2010**, *77*, 730-742.

33   van der Kamp, M. W.; Mulholland, A. J. Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology. *Biochemistry* **2013**, *52*, 2708-2728.

34   Singh, U. C.; Kollman, P. A. A Combined *Ab Iinitio* Quantum Mechanical and Molecular Mechanical Method for Carrying out Simulations on Complex Molecular Systems: Applications to the $CH_3Cl + Cl^-$ Exchange Reaction and Gas Phase Protonation of Polyethers. *J. Comput. Chem.* **1986**, *7*, 718-730.

[35]   Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103*, 227-249.

[36]   Field, M. J.; Bash, P. A.; Karplus, M. A Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations. *J. Comput. Chem.* **1990**, *11*, 700-733.

[37]   Thèry, V.; Rinaldi, D.; Rivail, J.-L.; Maigret, B.; Ferenczy, G. G. Quantum mechanical computations on very large molecular systems: The local self-consistent field method. *J. Comput. Chem.* **1994**, *15*, 269–282.

[38]   Ferenczy, Gy. G. Calculation of Wave-Functions with Frozen Orbitals in Mixed Quantum Mechanics/Molecular Mechanics Methods. Part I. Application of the Huzinaga Equation. *J. Comput. Chem.* **2013**, *34*, 854-861.

[39]   Ferenczy, Gy. G. Calculation of Wave-Functions with Frozen Orbitals in Mixed Quantum Mechanics/Molecular Mechanics Methods. Part II. Application of the Local Basis Equation. *J. Comput. Chem.* **2013**, *34*, 862-869.

[40]   Kästner, J.; Thiel, S.; Senn, H. M.; Sherwood, P.; Thiel, W. Exploiting QM/MM Capabilities in Geometry Optimization: A Microiterative Approach Using Electrostatic Embedding. *J. Chem. Theory Comput.* **2007**, *3*, 1064-1072.

41 Kumar, S.; Rosenberg, J. M.; Bouzida, D; Swendsen, R. H.; Kollman, P. A. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13*, 1011-1021.

42 Prasad, B. R.; Plotnikov, N. V.; Lameira, J.; Warshel, A. Quantitative Exploration of the Molecular Origin of the Activation of GTPase. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 20509-20514.

43 Plotnikov, N. V.; Prasad, B. R.; Chakrabarty, S.; Chu, Z. T.; Warshel, A. Quantifying the Mechanism of Phosphate Monoester Hydrolysis in Aqueous Solution by Evaluating the Relevant Ab Initio QM/MM Free-Energy Surfaces. *J. Phys. Chem. B* **2013**, *117*, 12807-12819.

44 Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562-12566.

45 Plotnikov, N. V.; Kamerlin, S. C.; Warshel, A. Paradynamics: An Effective and Reliable Model for Ab Initio QM/MM Free-Energy Calculations and Related Tasks. *J Phys Chem B* **2011**, *115*, 7950-7962.

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

47 Zheng, J. J.; Wang, D.; Thiel, W.; Shaik, S. QM/MM Study of Mechanisms for Compound I Formation in the Catalytic Cycle of Cytochrome P450cam. *J. Am. Chem. Soc.* **2006**, *128*, 13204-13215.

48   Brodie, H. J.; Kripalani, K. J.; Possanza, G. Mechanism of Estrogen Biosynthesis. VI. The Stereochemistry of Hydrogen Elimination at C-2 During Aromatization. *J. Am. Chem. Soc*. **1969**, *91*, 1241-1242.

49   Akhtar, M.; Calder, M. R.; Corina, D. L.; Wright, J. N. Mechanistic Studies on C-19 Demethylation in Oestrogen Biosynthesis. *Biochem. J*. **1982**, *201*, 569-580.

50   Groves, J. T.; Gross, Z.; Stern, M. K. Preparation and Reactivity of Oxoiron(IV) Porphyrins. *Inorg. Chem.* **1994**, *33*, 5065–5072.

51   He, K,; Falick, A. M.; Chen, B.; Nilsson, F.; Correia, M. A. Identification of the Heme Adduct and an Active Site Peptide Modified during Mechanism-Based Inactivation of Rat Liver Cytochrome P450 2B1 by Secobarbital. *Chem. Res. Toxicol.*, **1996**, *9*, 614–622.

52   de Visser, S. P.; Kumar, D.; Shaik, S. How do aldehyde side products occur during alkene epoxidation by cytochrome P450? Theory Reveals a State-Specific Multi-State Scenario where the High-Spin Component Leads to all Side Products. *J. Inorg. Biochem.* **2004**, *98*, 1183-1193.

53   Ogliaro, F.; Cohen, S.; de Visser, S. P.; Shaik, S. Medium Polarization and Hydrogen Bonding Effects on Compound I of Cytochrome P450: What Kind of a Radical Is It Really? *J. Am. Chem. Soc.* **2000**, *122*, 12892–12893.

54   Ogliaro, F.; de Visser, S. P.; Cohen, S.; Kaneti, J.; Shaik, S. The Experimentally Elusive Oxidant of Cytochrome P450: A Theoretical "Trapping" Defining More Closely the "Real" Species. *ChemBioChem* **2001**, *2, 848–851.*

55   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.* **2005**, *127*, 12900–12908.

56   Ogliaro, F.; Cohen, S.; Filatov, M.; Harris, N.; Shaik, S. The High-Valent Compound of Cytochrome P450: The Nature of the Fe−S Bond and the Role of the Thiolate Ligand as an Internal Electron Donor. *Angew. Chem., Int. Ed.* **2000**, *39*, 3851-3855.

57   Ogliaro, F.; Cohen, S.; Filatov, M.; Harris, N.; Shaik, S. The High-Valent Compound of Cytochrome P450: The Nature of the Fe−S Bond and the Role of the Thiolate Ligand as an Internal Electron Donor. *Angew. Chem., Int. Ed.* **2001**, *40*, 647.

58   Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A. et al. Gaussian 09, revision A.1; Gaussian, Inc.: Wallingford, CT, 2009.

59   Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648-5652.

60   Altun, A.; Kumar, D.; Neese, F.; Thiel, W. Multireference Ab Initio Quantum Mechanics/Molecular Mechanics Study on Intermediates in the Catalytic Cycle of Cytochrome P450$_{cam}$. *J. Phys. Chem. A* **2008**, *112*, 12904-12910.

61   Dolg, M.; Wedig, U.; Stoll, H.; Preuss, H. Energy-Adjusted Ab Initio Pseudopotentials for the First Row Transition Elements. *J. Chem. Phys.* **1987**, *86*, 866-872.

62   Andrae, D.; Haussermann, U.; Dolg, M.; Stoll, H.; Preuss, H. Energy-Adjusted Ab Initio Pseudopotentials for the Second and Third Row Transition Elements. *Theor. Chim. Acta* **1990**, 77, 123-141.

63   Krámos, B.; Menyhárd, D. K.; Oláh, J. Direct Hydride Shift Mechanism and Stereoselectivity of P450$_{nor}$ Confirmed by QM/MM Calculations. *J. Phys. Chem. B* **2012**, *116*, 872–885.

64   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.* **2010**, *132*, 154104.

65   NBO 5.9.; Glendening, E. D.; Badenhoop, J. K.; Reed, A. E.; Carpenter, J. E.; Bohmann, J. A.; Morales, C. M.; Weinhold, F. (Theoretical Chemistry Institute, University of Wisconsin, Madison, WI, 2009); http://www.chem.wisc.edu/~nbo5 (accessed Nov 4, 2013).

66   Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Functionals. *Theor. Chem. Acc.* **2008**, *120*, 215-241.

67   Tao, J. M.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the Density Functional Ladder: Nonempirical Meta-Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **2003**, *91*, 146401.

68   Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785-789.

69   Chai, J. D.; Head-Gordon, M. Long-range Corrected Hybrid Density Functionals with Damped Atom-Atom Dispersion Corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615-6620.

70   Miertuš, S.; Scrocco, E.; Tomasi, J. Electrostatic Interaction of a Solute with a Continuum. A Direct Utilization of ab initio Molecular Potentials for the Prevision of Solvent Effects. *Chem. Phys.* **1981**, *55*, 117-129.

71   Pascual-Ahuir, J. L.; Silla, E.; Tuñón, I. GEPOL: An Improved Description of Molecular-Surfaces. 3. A New Algorithm for the Computation of a Solvent-Excluding Surface. *J. Comp. Chem.* **1994**, *15*, 1127-1138.

72   Ghosh, D.; Griswold, J.; Erman, M.; Pangborn, W. Structural Basis for Androgen Specificity and Oestrogen Synthesis in Human Aromatase. *Nature* **2009**, *457*, 219-223.

73   Altun, A.; Guallar, V.; Friesner, R. A.; Shaik, S.; Thiel, W. The Effect of Heme Environment on the Hydrogen Abstraction Reaction of Camphor in P450$_{cam}$ Catalysis: A QM/MM Study. *J. Am. Chem. Soc.* **2006**, *128*, 3924–3925.

74   Altun, A.; Shaik, S.; Thiel, W. Systematic QM/MM Investigation of Factors that Affect the Cytochrome P450-Catalyzed Hydrogen Abstraction of Camphor. *J. Comput. Chem.* **2006**, *27*, 1324-1337.

[75] 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.* **2005**, *127*, 12900-12908.

[76] MacKerell, A. D, Jr.; Banavali, N.; Foloppe, N. Development and Current Status of the CHARMM Force Field for Nucleic Acids. *Biopolymers* **2000**, *56*, 257–265.

[77] Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. *J. Comput. Chem.* **2009**, *30*, 1545–1614.

[78] Harvey, J. N. Spin-Forbidden CO Ligand Recombination in Myoglobin. *Faraday Discuss.* **2004**, No. 127, 165–177.

[79] TINKER - Home Page. Tinker - Software tools for molecular design. http://dasher.wustl.edu/tinker/ (accessed October 5, 2011).

[80] Ren, P.; Wu, C.; Ponder, J. W. Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. *J. Chem. Theory Comput.* **2011**, *7*, 3143–3161.

[81] 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. Phys. Chem. Lett.* **2010**, *1*, 3232–3237.

[82] Lonsdale, R.; Harvey, J.N.; Mulholland, A.J., Effects of Dispersion in Density Functional Based Quantum Mechanical/Molecular Mechanical Calculations on Cytochrome P450 Catalyzed Reactions. *J. Chem. Theory Comput.* **2012**, *8*, 4637-4645.

83   Grimme, S. Semiempirical GGA-Type Density Functional Constructed with a Long-Range Dispersion Correction. *J. Comp. Chem.* **2006**, *27*, 1787-1799.

84   Schaftenaar, G.; Noordik, J. H. Molden: A Pre- and Post-Processing Program for Molecular and Electronic Structures. *J. Comput.-Aided Mol. Design* **2000**, *14*, 123-134.

85   Humphrey, W.; Dalke, A.; Schulten, K. VMD - Visual Molecular Dynamics *J. Molec. Graphics* **1996**, *14*, 33-38.

86   Schöneboom, J. C.; Cohen, S.; Lin, H.; Shaik, S.; Thiel, W. Quantum Mechanical/Molecular Mechanical Investigation of the Mechanism of C−H Hydroxylation of Camphor by Cytochrome P450cam: Theory Supports a Two-State Rebound Mechanism. *J. Am. Chem. Soc.* **2004**, *126*, 4017–4034.

87   Meigs, R. A.; Ryan, K. J. Enzymatic Aromatization of Steroids I. Effects of Oxygen and Carbon Monoxide on the Intermediate Steps of Estrogen Biosynthesis. *J. Biol. Chem.*, **1971**, *246*, 83-87.

88   Townsley, J. D.; Brodi, H. J. Mechanism of Estrogenbiosynthesis. III. Stereochemistry of Aromatization of C19 and C18 Steroids. *Biochemistry* **1968**, *7*, 33-40.

89   Conradie, J.; Abhik Ghosh, A. DFT Calculations on the Spin-Crossover Complex Fe(salen)(NO): A Quest for the Best Functional. *J. Phys. Chem. B* **2007**, *111*, 12621-12624.

90   Harvey, J. N. On the accuracy of density functional theory in transition metal chemistry. *Annu. Rep. Prog. Chem., Sect. C: Phys. Chem.* **2006**, *102*, 203-226.

91   Radoń, M. Revisiting the Role of Exact Exchange in DFT Spin-State Energetics of Transition Metal Complexes. *Phys Chem Chem Phys*. **2014**, *16*, 14479-14488.

92   Frushicheva, M. P.; Warshel, A. Towards Quantitative Computer-Aided Studies of Enzymatic Enantioselectivity: The Case of Candida antarctica Lipase A. *ChemBioChem* **2012**, *13*, 215−223.

93   Lonsdale, R.; Hoyle, S.; Grey, D. T.; Ridder, L.; Mulholland, A. J. Determinants of Reactivity and Selectivity in Soluble Epoxide Hydrolase from Quantum Mechanics/Molecular Mechanics Modeling. *Biochemistry* **2012**, *51*, 1774−1786.

94   Torrie, G. M.; Valleau, J. P. Monte Carlo Free Energy Estimates Using Non-Boltzmann Sampling: Application to the Sub-Critical Lennard-Jones Fluid. *Chem. Phys. Lett*. **1974**, *28*, 578-581.

95   Car, R.;Parrinello M. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.* **1985**, *55*, 2471-2474.

96   Ufimtsev, I. S.; Martínez T. J. Quantum Chemistry on Graphical Processing Units. 1. Strategies for Two-Electron Integral Evaluation. *J. Chem. Theory Comput.* **2008**, *4*, 222-231.

97   Sisto, A.; Glowacki, D. R.; Martinez, T. J. Ab Initio Nonadiabatic Dynamics of Multichromophore Complexes: A Scalable Graphical-Processing-Unit-Accelerated Exciton Framework. *Acc. Chem. Res.*, **2014**, *47*, 2857-2866.

98   Lonsdale, R.; Houghton, K. T.; Żurek, J.; Bathelt, C. M.; Foloppe, N.; de Groot, M. J.; Harvey, J. N.; Mulholland, A. J. Quantum Mechanics/Molecular Mechanics Modeling of Regioselectivity of Drug Metabolism in Cytochrome P450 2C9. *J. Am. Chem. Soc.* **2013**, *135*, 8001−8015.

99   Oláh, 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. U.S.A.* **2011**, *108*, 6050−6055.

100   Ranaghan, K. E.; Mulholland, A. J. Investigations of Enzyme-Catalysed Reactions with Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods. *Int. Rev. Phys. Chem.* **2010**, *29*, 65-133.

101   Klähn, M.; Braun-Sand, S.; Rosta, E.; Warshel, A. On Possible Pitfalls in ab Initio Quantum Mechanics/Molecular Mechanics Minimization Approaches for Studies of Enzymatic Reactions. *J. Phys. Chem. B* **2005**, *109,* 15645-15650

102   Harvey, J. N.; Bathelt, C. M.; Mulholland, A. J. QM/MM Modeling of Compound I Active Species in Cytochrome P450, Cytochrome C Peroxidase, and Ascorbate Peroxidase. *J. Comput. Chem.* **2006**, *27*, 1352-1362.

103   Warshel, A.; Sharma, P. K.; Kato, M.; Xiang Y.; Liu H.; Olsson, M. H. Electrostatic Basis for Enzyme Catalysis. *Chem Rev.* **2006**, *106*, 3210-3235.

104   Warshel, A. Energetics of Enzyme Catalysis. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 5250-5254.