J|A|C|S
COMMUNICATIONS

Published on Web 08/03/2006

Does Substrate Oxidation Determine the Regioselectivity of Cyclohexene and Propene Oxidation by Cytochrome P450?

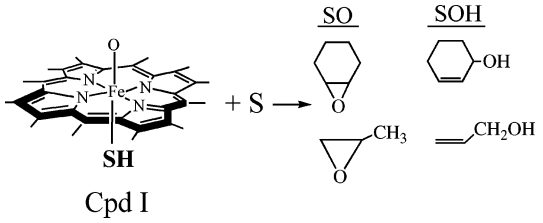
Shimrit Cohen, Sebastian Kozuch, Carina Hazan, and Sason Shaik*

Department of Chemistry and the Lise Meitner-Minerva Center for Computational Quantum Chemistry, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel

Received May 10, 2006; E-mail: sason@yfaat.ch.huji.ac.il

The nature of the factors that govern the oxidative regiochemistry by the enzyme cytochrome P450 is a fundamental problem. Two of the key reactions of P450s are C−H hydroxylation and C═C epoxidation.[1−5] Different isozymes give different ratios of these products with substrates capable of undergoing both reactions. For example, with cyclohexene, the ratios C−H/C═C are 0.5−1, using $P450_{cam}$, its mutants, and $P450_{LM2}$.[5a,b] By contrast, propene, 1-butene, and 1-octene give exclusively epoxidation-type products with $P450_{LM2}$,[3,5c] while 2-butenes give ratios of 0.023−0.025,[5d] and unsaturated fatty acids give C−H/C═C ratios larger than 1 with $P450_{BM3}$.[5e] The present paper addresses the regiochemical problem using the reactions of cyclohexene and propene (**S**) with the active species, Compound I (Cpd I), of a single P450 isozyme, $P450_{cam}$; see Scheme 1. The study uses density functional theory (DFT) and

Scheme 1. Hydroxylation and Epoxidation of Cyclohexene and Propene by Cpd I



considers the rate-controlling step of the oxidation process, itself, as well as the overall kinetics of the entire catalytic cycle.[6] The selectivity was studied therefore at three levels: (i) gas-phase calculations were carried out to reveal intrinsic selectivity trends; (ii) hybrid QM/MM calculations were performed to examine the impact of the protein environment on the regiochemistry; and (iii) the regioselectivity was calculated using the effective barrier of a cycle under steady-state conditions, where steps other than those of the oxidative process affect the turnover frequency of the cycle.[6] As shall be demonstrated, the C−H/C═C ratio is entirely different under turnover conditions and in noncatalytic processes.

Both gas-phase[7a] and QM/MM[7b] calculations were carried out using UB3LYP and the LACVP(Fe)/6-31G(C,H,N,O) basis set, B1, which was used to optimize geometries. For the gas-phase calculations, the critical points were ascertained by their frequencies; the corresponding ZPE corrections were used for all other calculations. Energies were corrected with the more extended basis sets: B2, B2W, B3, and B4, where W is the all-electron Wachters basis set,[7c] augmented with diffuse and polarization functions on Fe, while B2 employs 6-31G*(C,H,N,O) on the O,N,S ligands to iron, B3 uses 6-31+G* on these ligands, and B4 uses 6-31+G* also on the substrate. The corrected energies are abbreviated as B*n*//B1 (n = 2, 2W, 3, or 4). The QM/MM calculations utilized the system set up described earlier for the reactions of $P450_{cam}$.[8] The QM subsystem included Cpd I, oxo-iron porphine with an axial SH

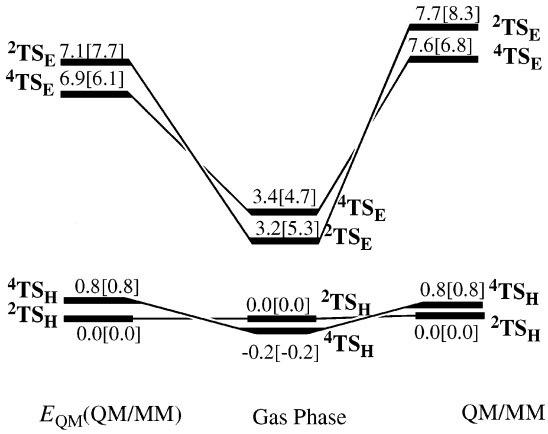


Figure 1. Relative TS energies for bond activation during the oxidation of cyclohexene. In the middle are gas-phase data. On the right we show QM/MM energies (averaged over three snapshots), and on the left the QM part of the QM/MM energies. All values include ZPE correction. Figures out of brackets correspond to B1, while in brackets to B2W//B1.

ligand, and the substrates in Scheme 1. In both reactions, we selected from the MD equilibration trajectory a few snapshots to perform the QM/MM optimization. The results of the different snapshots were similar, and hence we describe here representative key data. The data are summarized in the Supporting Information.

Figure 1 shows the relative energies of the quartet and doublet transition states[9] for the bond activation steps in hydroxylation ($^{2,4}TS_H$) and epoxidation ($^{2,4}TS_E$) reactions of Cpd I with cyclohexene. In the gas phase, the preference for hydroxylation is determined primarily by the ZPE correction, which is large for the hydroxylation and negligible for epoxidation. In the protein pocket, the preference for hydroxylation becomes larger, reaching ca. 7−8 kcal/mol on average. It is seen that this preference is determined by the QM part of the QM/MM energy, hence, by the electrostatic effect of the protein. Similar data for propene oxidation are summarized in the Supporting Information (Tables S41−S50, Figures S7−S10); in the gas phase, epoxidation is preferred by ca. 0.3 kcal/mol,[9] but the protein establishes a preference for hydroxylation over epoxidation of ca. up to 5.5 kcal/mol depending on the snapshot. Thus, for the two substrates, *the electrostatic and hydrogen bonding machineries of the protein pocket create preference for allylic hydroxylation*, as predicted before using simple model systems.[9]

These conclusions focus on the oxidation process and do not take into account the entire catalytic cycle. Indeed, the predictions in Figure 1 do not match the available experimental data for the two substrates.[3,5a−c] As such, we turn to consider the selectivity question based on a model developed for catalytic cycles.[6] The free energy of activation, δE, of a cycle in steady state is given approximately by the maximum energy difference between a

10.1021/ja063269c CCC: $33.50

specific pair of a TS and an intermediate species, which we refer to here as the rate-determining TS (RDTS) and the most abundant reaction intermediate (MARI). Whenever the RDTS precedes the MARI, the effective activation energy of the cycle, δE, will include also the reaction free energy, ΔE:

$$\delta E = E_{\mathrm{RDTS}} - E_{\mathrm{M}} + \Delta E \quad (1)$$

In $P450_{cam}$, the RDTS of the cycle is the one for the second electron transfer.[4,10] Furthermore, as can be deduced from DFT[11] and QM/MM calculations,[8] the MARI of the cycle is the ferric alcohol for hydroxylation and the ferric epoxide for epoxidation. *Assuming fast mobility[12] between the C═C and C−H sites* of the substrate and having a common RDTS for both processes, the regiochemistry will be determined by the following difference between the effective barriers:

$$\delta E_{\mathrm{H}} - \delta E_{\mathrm{E}} = \Delta\delta E_{\mathrm{HE}} = E_{\mathrm{M,E}} - E_{\mathrm{M,H}} + \Delta E_{\mathrm{H}} - \Delta E_{\mathrm{E}} \quad (2)$$

This barrier difference is given by the energy difference of the MARI species (E_{M}) of epoxidation (E) and hydroxylation (H) and the corresponding difference between the reaction energies for the net processes (ΔE_{H} and ΔE_{E}). Since the net reaction involves common reactants, the substrate (S) and $O_2 + 2e^- + 2H^+$,[4] the $\Delta E_{\mathrm{H}} - \Delta E_{\mathrm{E}}$ term becomes the difference between the energies of the free epoxide (SO) and alcohol (SOH), and eq 2 is simplified to

$$\Delta\delta E_{\mathrm{HE}} = E_{\mathrm{M,E}} - E_{\mathrm{M,H}} + E_{\mathrm{SOH}} - E_{\mathrm{SO}} \quad (3)$$

Thus, under free mobility between the oxidation sites, the difference of the activation barriers will be determined by the relative ease of release of the two products from their MARIs. This does not mean that the product release is rate determining, only that regiochemistry follows the energetic of product release.

Since free energies are not available, eq 3 is used with energies corrected by ZPE.[6] In addition, the energy difference of the free oxidation products ($\Delta\Delta E = E_{\mathrm{SOH}} - E_{\mathrm{SO}}$) corresponds to conditions "outside" of the active site on the protein surface; this can be taken as either an aqueous[8] or a gas-phase environment.

Equation 3 leads to a clear-cut conclusion, which is independent of snapshot, basis set, or the polarity of the environment of the free products (Tables S3−S15). Thus, using the lowest MARI at the best level (B4//B1), the $\Delta\delta E_{\mathrm{HE}}$ values for cyclohexene are positive, in the range of 0.39−2.63 kcal/mol, while for propene it is 5.68−11.93 kcal/mol. Using these $\Delta\delta E_{\mathrm{HE}}$ values in the Eyring equation, at room temperature, leads to C−H/C═C ratios of 0.52−0.012 for cyclohexene and much lower ones, ca. 7×10^{-5} to 7×10^{-9}, for propene. Thus, the preferred regiochemistry predicted by eq 3 for both substrates is C═C epoxidation. This is in accord with experiment;[3,5a−c] the quantitative preference is small in cyclohexene, for which experimental C−H/C═C values are 0.5,[5a] while the result for propene matches the finding of exclusive epoxidation products.[5c]

The root causes of this uniform prediction of eq 3 are straightforward: the alcohol binds to the heme more strongly than the epoxide, and therefore, the MARI for C−H hydroxylation is lower in energy than the corresponding species in C═C epoxidation. This energy difference dominates the effective activation energies of the cycles and overrides the opposing difference in the relative stabilities of the free alcohol and epoxide products. *As such, the $\Delta\delta E_{HE}$ quantity in eq 3 is positive, thereby preferring C═C epoxidation*. The actual value of $\Delta\delta E_{\mathrm{HE}}$ is affected by the substrate and the interactions inside the protein pocket, wherein the ferric alcohol complex is more tightly bonded than the ferric epoxide complex. Cyclohexenol and cyclohexene oxide are more weakly bound to the heme than the corresponding products of propene, and hence the $\Delta\delta E_{\mathrm{HE}}$ quantity is small for cyclohexene and significant for propene. While proper sampling of the conformations is needed for good quantitative predictions, this will not affect the conclusion that, under steady-state conditions, $P450_{cam}$ will perform oxidation of propene and cyclohexene with preference for C═C epoxidation. However, under conditions of a single turnover (or when Cpd I is prepared and reacts with the substrate in a noncatalytic manner), the regiochemistry will be determined by the relative energies of the TSs for the two competing processes (Figure 1); under these conditions, the enzyme is predicted to prefer C−H hydroxylation by a wide margin augmented by the protein environment effect.[9]

Thus, assuming free mobility between the oxidation sites,[12] the following additional predictions can be made. (a) Whenever the RDTS in the cycle is not a step of the oxidation process, a preference for allylic C−H hydroxylation would be expected only when the ferric alcohol MARI is destabilized relative to the ferric epoxide. (b) Whenever the RDTS *does involve* substrate oxidation,[10] the C−H/C═C ratio will reflect also the relative TS energies of the oxidation processes.[6] If the TSs behave as in Figure 1, such enzymes will exhibit greater propensity toward C−H hydroxylation than those in (a), and generally, regiochemistry in these enzymes will follow expected substituent effects on TSs. (c) In enzymes with several RDTSs and MARIs, of similar energies, the C−H/C═C ratio for a given substrate will be determined by the relative energies of the corresponding MARIs and RDTSs species.

Acknowledgment. The research is supported by a German Israeli Project (DIP) grant.

Supporting Information Available: Additional tables (54), figures (11), Cartesian coordinate, and full refs 7a,b. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Ortiz de Montellano, P. R., Ed. *Cytochrome P450: Structure, Mechanism and Biochemistry*, 2nd ed.; Plenum Press: New York, 1995. (b) Ortiz de Montellano, P. R., Ed. *Cytochrome P450: Structure, Mechanism and Biochemistry*, 3rd ed.; Kluwer Academic/Plenum Publishers: New York, 2005.
(2) (a) Ortiz de Montellano, P. R.; De Voss, J. J. *Cytochrome P450: Structure, Mechanism and Biochemistry*, 2nd ed.; Plenum Press: New York, 1995; p 245. (b) Ortiz de Montellano, P. R.; De Voss, J. J. *Cytochrome P450: Structure, Mechanism and Biochemistry*; Kluwer Academic/Plenum Publishers: New York, 2005; p 183.
(3) Groves, J. T.; Han, Y.-Z. In *Cytochrome P450: Structure, Mechanism and Biochemistry*, 2nd ed.; Ortiz de Montellano, P. R., Ed.; Plenum Press: New York, 1995; p 1.
(4) Sono, M.; Roach, M. P.; Coulter, E. D.; Dawson, J. H. *Chem. Rev.* **1996**, *96*, 2841.
(5) (a) Yoshioka, S.; Takahashi, S.; Ishimori, K.; Morishima, I. *J. Inorg. Biochem.* **2000**, *81*, 141. (b) White, R. E.; Groves, J. T.; McClusky, G. A. *Acta Biol. Med. Germ.* **1979**, *38*, 475. (c) Groves, J. T.; Avaria-Neisser, G. E.; Fish, K. M.; Imachi, M.; Kuczkowski, R. L. *J. Am. Chem. Soc.* **1986**, *108*, 3837. (d) Vaz, A. D. N.; McGinnity, D. F.; Coon, M. J. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 3555. (e) Ruettinger, R. T.; Fulco, A. J. *J. Biol. Chem.* **1981**, *256*, 5728.
(6) Kozuch, S.; Shaik, S. *J. Am. Chem. Soc.* **2006**, *128*, 3355.
(7) (a) Ahlrichs, R. *TURBOMOLE 5.5*, University of Karlsruhe. (b) Sherwood, P.; et al. *J. Mol. Struct.* **2003**, *632*, 1. (c) Wachters, A. J. H. *J. Chem. Phys.* **1970**, *52*, 1033.
(8) (a) Schöneboom, J. C.; Cohen, S.; Lin, H.; Shaik, S.; Thiel, W. *J. Am. Chem. Soc.* **2004**, *126*, 4017. (b) Schöneboom, J. C.; Thiel, W. *J. Phys. Chem. B* **2004**, *108*, 7468. (c) Lin, H.; Schöneboom, J. C.; Cohen, S.; Theil, W. *J. Phys. Chem. B* **2004**, *108*, 10083.
(9) De Visser, S. P.; Ogliaro, F.; Sharma, P. K.; Shaik, S. *J. Am. Chem. Soc.* **2002**, *124*, 11809.
(10) See however: Yun, C.-H.; Miller, G. P.; Guengerich, F. P. *Biochemistry* **2001**, *40*, 4521.
(11) (a) Loew, G. H.; Harris, D. L. *Chem. Rev.* **2000**, *100*, 407. (b) Kamachi, T.; Yoshizawa, K. *J. Am. Chem. Soc.* **2003**, *125*, 4652.
(12) Volz, T. J.; Rock, D. A.; Jones, J. P. *J. Am. Chem. Soc.* **2002**, *124*, 9724.

JA063269C