

Supporting Information

- ✧ Details of the Simulations
- ✧ Validation of Density Functional and Determining the Ground State
- ✧ Ab initio QM/MM Molecular Dynamics (MD) Simulation
- ✧ **Table S1.** The total energy difference for the singlet, triplet and quintet states in the cluster model
- ✧ **Table S2.** The optimized geometric parameters of ferric-superoxide species
- ✧ **Table S3.** The relative energies estimated with five types of functionals in the first proton transfer process.
- ✧ **Table S4.** The optimized geometric parameters of bond lengths (in Å) and angles (in degree) in QM/MM MD
- ✧ **Figure S1.** Evolutions of mean square displacements of water and protein along with time
- ✧ **Figure S2.** Three-dimensional density distribution of water around the iron atom
- ✧ **Figure S3.** Optimized structures of transition states and intermediates in the proximal / distal proton transfer reaction.
- ✧ **Figure S4.** QM/MM optimized structures of transition state and intermediates are shown about the fate of the hydroperoxo species.
- ✧ **Figure S5.** The optimized structures of transition state and intermediates in the carbon-carbon bond cleavage reaction pathway.

- ✧ **Figure S6.** The optimized structures of several key species in the reaction pathways
- ✧ **Figure S7.** The total electron density differences of species **4a** and **4b** related to the carbon-carbon bond cleavage step.
- ✧ **Scheme 1.** Suggested mechanism in the catalytic cycle of HEPD

Details of Simulations

The theoretical model (Figure 1a) was set up based on the X-ray crystal structure of HEPD in complexation with the substrate 2-HEP (PDB ID: 3GBF).^[1] The missing residues, i.e. 38-45, 297-299, and 342-348, were refined by Modeller.^[2] The mutated residues (selenomethionines) were changed back to methionines, and the coordinated crystal water W_{725} was replaced by the oxygen molecule O_2 . The coordinated chromium (Cd^{II}) in the active site was manually changed to iron (Fe^{II}), while the others were removed. The protonation states of ionizable amino acids were selected and checked based on a visual inspection of their microenvironments by VMD program.^[3] In the active site, the protonation states of His₁₂₉ and His₁₈₂ were kept in neutral state, and the Glu₁₇₆ was deprotonated. The hydrogen atoms of HEPD were added by HBUILD module in CHARMM program.^[4] The resulting system was solvated with a water box, and this solvated model of HEPD contained about 40000 atoms, including about 6700 protein atoms.

The prepared structure was first minimized by CHARMM program with the CHARMM22 force field^[5] for protein and TIP3P model^[6] for water. The classical molecular dynamic (MD) simulation at the force field level was performed and extended to 20 ns under stochastic boundary conditions^[7] in the canonical (NVT) ensemble. Mean-square displacement (MSD) was adopted to depict the diffusive behavior of the system. In this MD simulation, the temperature was maintained at 300 K. The residues coordinated with iron were kept fixed during these simulations, since it is very difficult to properly describe transition metal coordination shell at the molecular mechanical force field level. The MSDs of the protein and water atoms (Figure S1) relative to the initial structure become linear within 3 ns, and the system is considered in its equilibrium states. The final simulated structure has a little larger deviation from the crystal structure since the unnecessary Cd^{II} was removed in our model. The root-mean squared deviations (RMSD) between the crystal structure and the final model structure is nearly 2.1 Å.

Then the snapshot of the equilibrated MD trajectory was taken as the starting point for further QM and QM/MM simulations. Our QM/MM model of HEPD contains 32734 atoms, including 6671 protein atoms and 8681 water molecules. The 2-His-1-Glu facial triad (His₁₂₉, His₁₈₂, Glu₁₇₆), HEP, O₂, and two water molecules are included in the QM region. During the QM/MM geometry optimizations, the QM region and 7813 MM atoms (defined by including the residues within 20Å of the QM region) were allowed to move, whereas all the remaining atoms were kept fixed as the environment. The frequency was obtained by a two-point displacement method with

analytical calculation of gradients, and electronic polarization of the environment was included in the frequency calculation.

Validation of Density Functional and Determining the Ground State

As discussed in some bioinorganic interesting systems,^[8] it is important to consider the ground state in DFT calculations for a transition metal containing biological molecule. Therefore the ground state and functionals used in our model were carefully validated with high level *ab initio* calculations. Three spin states (singlet, triplet and quintet) of the active site in HEPD were considered in our studies. The cluster model of active site was prepared as mentioned above, and comprised 52 atoms, with a total charge of -1. The procedure of our calculations is similar as the studies of many iron complexes.^[9] The other calculations were performed as implemented in the TURBOMOLE program package.^[10] Because the size of the bioinorganic species was too large for high level post-HF calculations (i.e. more than 1000 basis functions), our conclusion was only meaningful for qualitative comparison.

This cluster model was first optimized for the singlet state in gas phase at RIMP2/def2-TZVPP level.^[11] Then the optimized structure was employed to perform calculations for the three states with the second-order approximate coupled cluster method (RI-CC2) and def2-TZVPP basis sets^[12] as implemented in the ricc2 model of TURBOMOLE.^[11a, 13] In these calculations, 1248 basis functions (AOs) were included. The resolution-of-the-identity (RI) approximation was considered because it is more efficient than its conventional counterpart (i.e. 5 times faster, with

insignificant error) for this system with more than 1000 basis functions.^[14] To our best knowledge, the coupled cluster method is more stable and reliable than most single reference correlation methods.^[15] However, accurate results can only be obtained in conjunction with large basis sets. It perhaps not very meaningful to perform a CCSD(T) calculation with a double-zeta basis set. In order to obtain meaningful results, the least basis set like def2-TZVP(-f) or def2-TZVPP (cc-pVTZ, ano-pVTZ) would be required. Thus, our results could only give qualitative conclusions. For accurate results, quadruple-zeta and even larger basis sets are required. At this stage, the computational effort for these methods is too high (about 2000 basis functions).

In conclusion, the RI-CC2 results suggest that the single state is the ground state. The triplet and quintet states were 121.3 and 128.4 kcal/mol higher than the singlet state. The MP2 method also indicates that the singlet state is the ground state. And the triplet and quintet states are 71.0 and 44.7 kcal/mol higher than the singlet state. The value of energy difference seems to be too large for such kind of species. Usually, the energy difference of such non-heme species were much smaller, such as the case of catechol dioxygenase and cysteine dioxygenase.^[16] This result can be interpreted in three respects: (1) The RI-CC2 method is not as stable as CCSD(T) for ground state. (2) The basis set quality in our calculations was only considered to give qualitative results; (2) The single reference wavefunction may not be sufficient to describe this system precisely, as suggested in the studies of P450 and catechol dioxygenase.^[9b, 17] After all, it is still an open issue for which spin state is ground state.

Single point calculations were then performed for singlet/triplet/quintet states with def2-TZVPP basis set and 11 approximate functionals, including 5 GGA functionals (BLYP, BVWN, B-P, PBE, B97-D), 5 hybrid functionals (B3LYP, BHLYP, PBE0, TPSSH, B2PLYP) and one meta-GGA, TPSS. Then optimization was performed to determine the key geometry parameters for the singlet state with B3LYP(III), PBE0, B3LYP(V), TPSSH, PBE, B-P BLYP, TPSS, B97-D, and HF method with B1 basis set. Table S1 gives energy differences calculated by different functionals. The same basis set (def2-TZVPP) as used in RIMP2 and RI-CC2 calculations was adopted (1248 basis functions). The Hartree-Fock (HF) method provide a totally reversed stability order and the quintet state is predicted as the ground state. And triplet and quintet states are 112.5 and 167.8 kcal/mol lower than the singlet state. We can find that, for all GGA functionals, a lower total energy for the singlet spin state is predicted as obtained in the RI-CC2 calculations. For all hybrid functionals, as the weight of exact HF exchange increases, the results are approaching the HF results (Table S1). In additional calculations, we also find that the basis sets and the protein environment show insignificant effects on the conclusion.

Several optimized geometry parameters using various density functionals are given in Table S2. The average values over these parameters are calculated, and the standard deviation is adopted to compare the quality of the geometry structures. Standard deviations of the optimized geometry parameters from different functionals are defined as below, and it loops over all bond lengths and angle:

$$Dev\% = \sqrt{\sum_i [(X_i - Avg_i) / Avg_i]^2} \times 100\%$$

We can observe that PBE and B-P give the best balance among these functionals. And the deviations obtained from hybrid functionals are relatively larger, except the TPSSH functional which is based on a meta-GGA, TPSS. Another interesting result is the optimized parameters from HF calculations, e.g. the O-O bond distance of O₂ is about 1.16 Å, which is similar to the bond distance of free O₂ molecule. Thus the predicted Fe-O bond distance is also too long in HF calculations, and the HF fails to predict the interaction between O₂ and iron.

We notice that the GGA forms using PBE, B-P functionals and the meta-GGA (TPSS) show the balance in energy and optimization calculations, and it's also comparable with MP2 results (Table S4). In this work, PBE was chosen to perform optimization for the system. However, since there is no absolutely convincing reason to prefer this functional to the others, single point calculations were also performed with using B-P, B-LYP, PBE0 and B3LYP functionals^[18] as implemented in TURBOMOLE when the optimizations were done with PBE.

In conclusion, spin state energies of iron complexes are important for biochemical applications in order to produce more reliable results. Correlation plays an important role in the energy gaps of high spin and low spin states, high level of theoretical method for prediction energy gaps of low spin and high spin states still in need.

Ab initio QM/MM Molecular Dynamics (MD) Simulation

QM (PBE)/MM molecular dynamics (MD) simulation was performed in the NVT ensemble. The temperature was maintained at 300 K with standard Berendsen thermostat (characteristic time $\tau = 0.1$ ps). The time step was set to 0.001 ps. Our QM/MM model was first equilibrated for 3 ps to reach converged properties, and later the simulation was last for 10 ps.

Table S1. The total energy differences for singlet, triplet and quintet states in the cluster model of HEPD obtained from various density functionals with def2-TZVPP basis set. (GGA: the Generalized Gradient Approximation; HYB: hybrid functionals; HF: Hartree-Fock method)

methods	descriptions	singlet	triplet	quintet
B-LYP	GGA	0.0	3.1	20.9
B-VWN	GGA	0.0	3.3	21.6
PBE	GGA	0.0	2.5	17.4
B-P	GGA	0.0	2.4	17.7
TPSS	meta-GGA	0.0	1.6	18.8
B97-D	GGA	0.0	0.4	5.4
TPSSH	HYB 0.1*HF	0.0	-3.1	9.2
B3-LYP (V)	HYB 0.2*HF	0.0	-6.2	1.2
B3-LYP (III)	HYB 0.2*HF	0.0	-6.4	0.6

PBE0	HYB	0.25*HF	0.0	-10.6	-12.1
BH-LYP	HYB	0.5*HF	0.0	-38.0	-27.6
B2-LYP	DHYB	0.53*HF	0.0	-42.7	-34.3
HF		1.0*HF	0.0	-112.5	-167.8

Table S2. The optimized geometric parameters such as bond lengths (in Å) and angles (in degree) of the Fe^{III}-[O₂⁻] species are shown. They are optimized by DFT/B1, and various density functionals are used. Average values and standard deviations about different functionals are also shown.

<i>bond lengths</i> <i>/ angles</i>	<i>B3LYP</i> <i>(III)</i>	<i>B3LYP</i> <i>(V)</i>	<i>PBE0</i>	<i>TPSSH</i>	<i>PBE</i>	<i>B-P</i>	<i>B-LYP</i>	<i>TPSS</i>	<i>B97-D</i>	<i>*Avg</i>	<i>HF</i>
Fe - O _p	1.909	1.907	1.797	1.731	1.745	1.745	1.761	1.743	1.743	1.787	2.437
Fe - O ₂ _{HEP}	1.973	1.972	1.963	1.954	1.971	1.973	2.001	1.964	2.006	1.975	2.050
Fe - O ₄ _{HEP}	1.846	1.844	1.861	1.860	1.881	1.877	1.886	1.868	1.867	1.866	1.978
Fe - N _{His129}	2.056	2.055	2.019	2.008	2.008	2.008	2.051	2.007	2.054	2.030	2.150
Fe - N _{His182}	2.083	2.083	2.072	2.089	2.123	2.120	2.173	2.104	2.215	2.118	2.137
Fe - OE1 _{Glu176}	2.012	2.013	1.999	2.002	2.021	2.023	2.053	2.007	2.061	2.021	2.107
O _p - O _d	1.310	1.310	1.272	1.313	1.322	1.326	1.340	1.330	1.313	1.315	1.163
Fe-O _p -O _d	119.4	119.6	129.3	122.8	123.1	123.3	123.9	122.7	124.3	123.2	121.8
Dev %	2.78	2.73	2.31	1.37	0.98	0.99	1.55	1.11	2.11	/	/

Table S3. The relative energies estimated with five types of functionals in the first proton transfer process (distal (**D**) and proximal (**P**)) are listed for reference. And the results obtained in gas phase calculation are also shown. Relative energies (kcal/mol)

	QM/MM					QM
	GGA		HYB			PBE
	PBE	B-P	BLYP	PBE0	B3LYP	
R(P)	0.0	0.0	0.0	0.0	0.0	0.0
T(P)	17.8	17.3	19.7	9.3	12.7	21.8
P(P)	-2.2	-2.9	-2.2	-28.8	-23.6	9.8
R(D)	3.8	3.9	4.5	3.9	4.5	1.9
T(D)	25	25.3	27.5	18.8	22.4	21.2
P(D)	1.7	2	3.7	-20.3	-14.4	-3.6

Table S4. The optimized geometric parameters of bond lengths (in Å) and angles (in degree) of the Fe^{III}-[O₂⁻] species in gas phase and QM/MM model. The average value of ab initio QM/MM MD calculations was shown (the fluctuant range was also given). The geometric parameters of MP2/def2-TZVPP optimization were given for reference. Adopted nomenclature for the atoms was based on the name in crystal structure (PDB ID: 3GBF).

<i>bond lengths (Å)</i> <i>/ angles (degree)</i>	<i>PBE/MM</i>	<i>PBE</i>	<i>QM/MM MD</i> <i>(10ps)</i>	<i>MP2</i>
Fe - O _p	1.750	1.745	1.746±0.10	1.728
Fe - O _{2HEP}	2.030	1.971	2.058 ±0.20	1.870
Fe - O _{4HEP}	1.836	1.881	1.861±0.13	1.766
Fe - N _{His129}	2.061	2.008	2.056±0.15	1.943
Fe - N _{His182}	2.098	2.123	2.155±0.22	1.978
Fe - O _{Glu176}	2.000	2.021	2.029±0.20	1.878
O _p - O _d	1.335	1.322	1.320±0.05	1.370
CA _{HEP} - CB _{HEP}	1.554	1.545	1.554±0.08	1.529
Fe-O _p -O _d	121.0	123.1	120.3±10	116.9

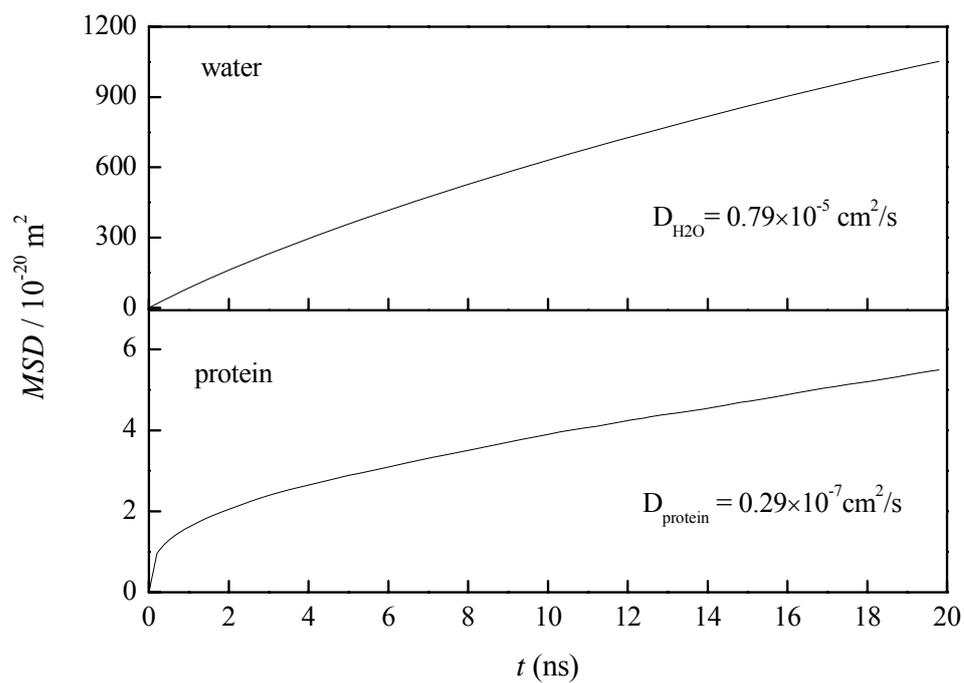


Figure S1. Evolutions of mean square displacements of water and protein along with time;

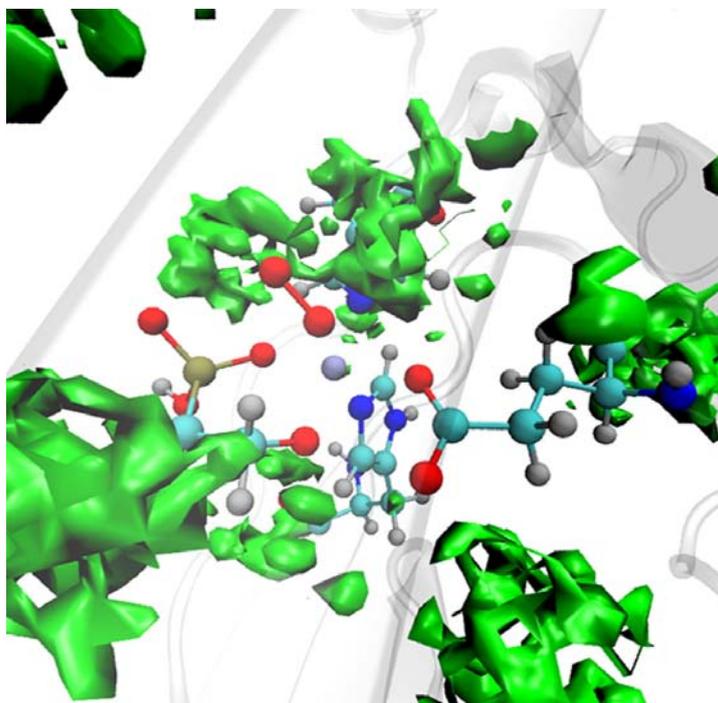


Figure S2. Three-dimensional density distribution (SDF) of water around the iron atom is given. The distance range of 0-1.0 nm is considered in the calculation of three-dimensional density distribution.

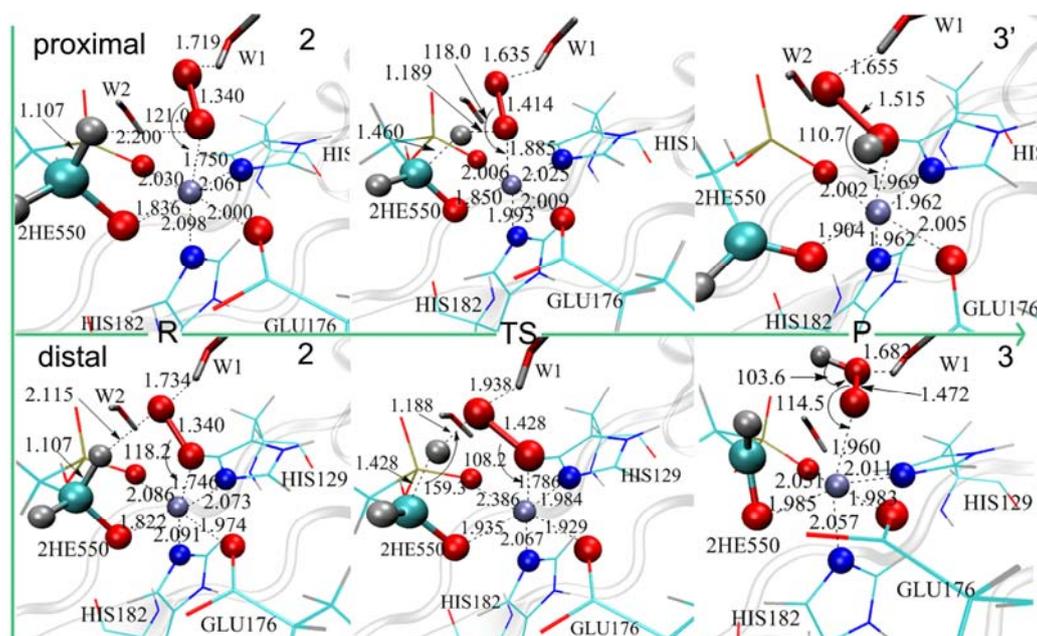


Figure S3. Optimized structures of transition states and intermediates in the proximal / distal proton transfer reaction. The coordinated atoms and other relevant atoms in the reaction are highlighted for clarity. The species **2** in the proximal and distal process are similar with slightly different configurations.

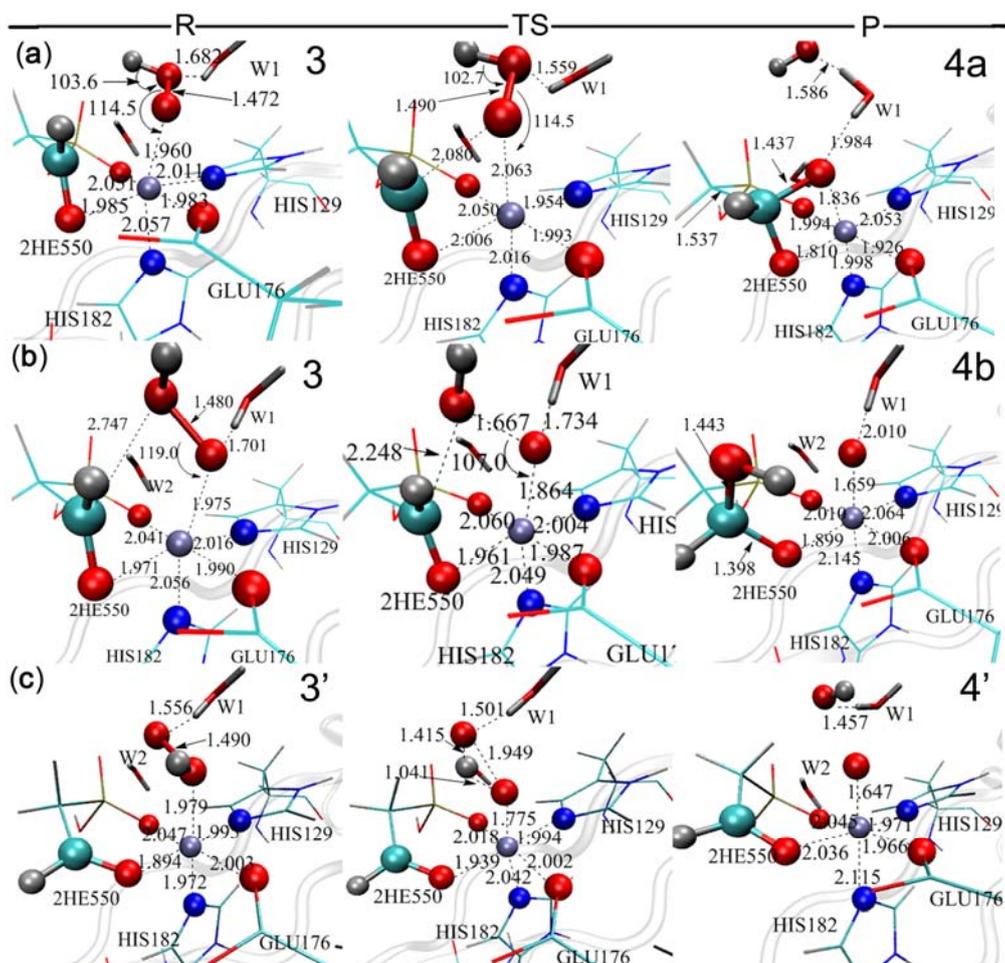


Figure S4. QM/MM optimized structures of transition state and intermediates are shown about the fate of the hydroperoxo species. The species **3** in **(a)** and **(b)** are similar with different configurations.

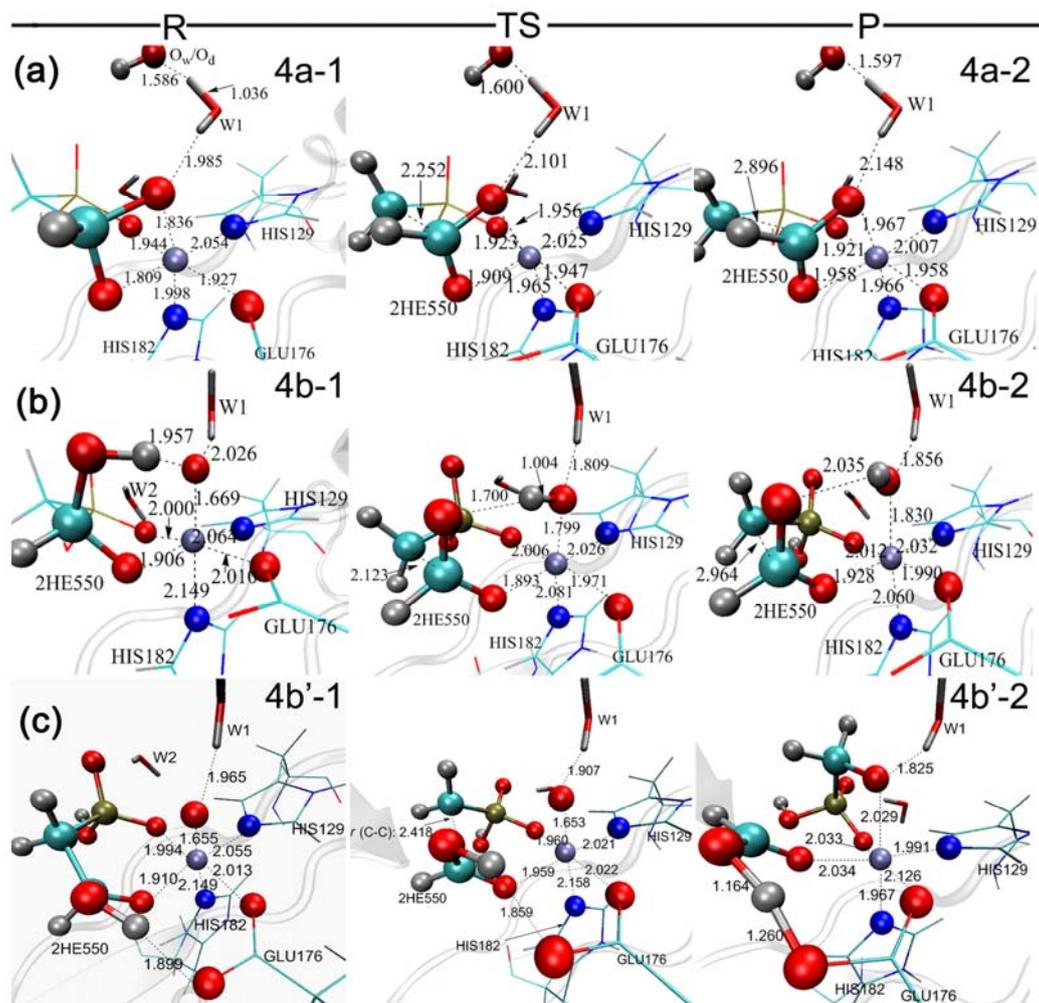


Figure S5. The optimized structures of transition state and intermediates in the carbon-carbon bond cleavage reaction pathway. (a) mechanism a, (b) mechanism b and (c) mechanism b'. The species **4b-1** and **4b'-1** are similar complexes with different configurations.

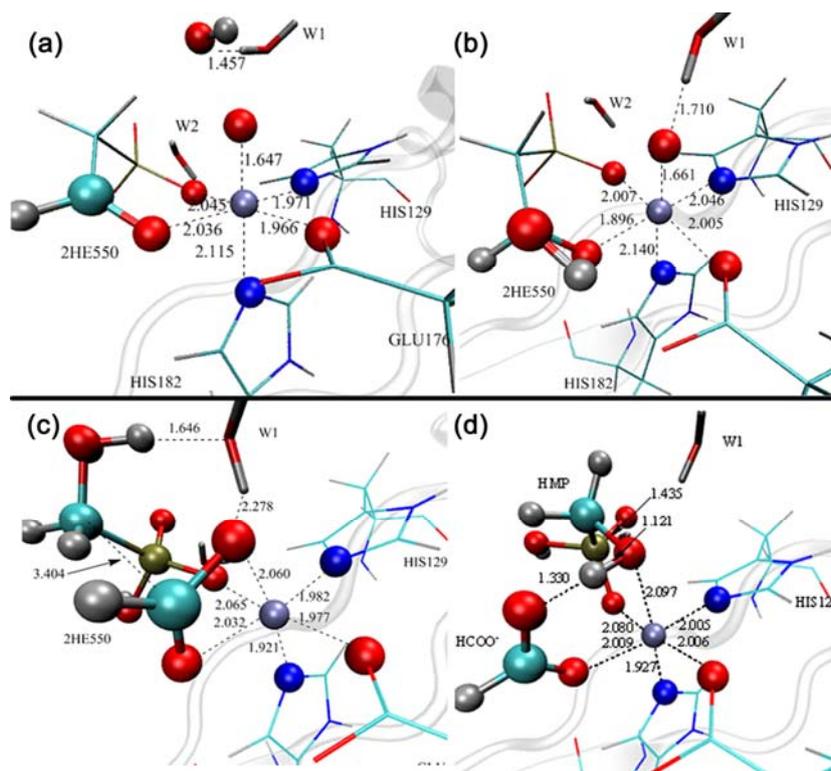


Figure S6. The optimized structures of several key species in the reaction pathways are shown. The intermediates after the proximal proton transfer step are given in the upper panel (**a**, **b**). And the intermediates after the carbon-carbon bond cleavage step (path **a**, **b**) are given in the bottom panel (**c** and **d**). (Colors: iceblue for the iron atom, blue for nitrogen atom, cyan for carbon atom, brown for the phosphorus atom, and gray for hydrogen atom)

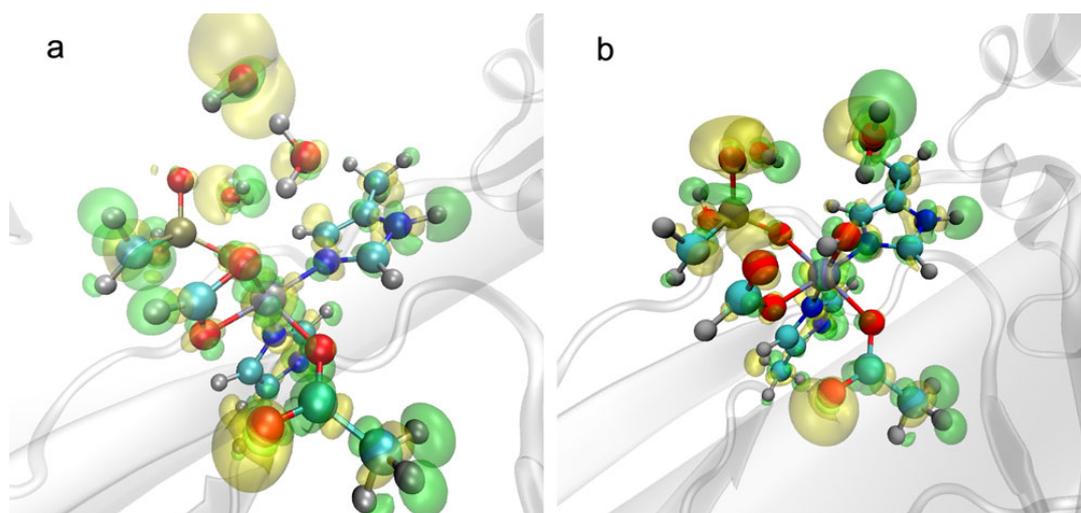
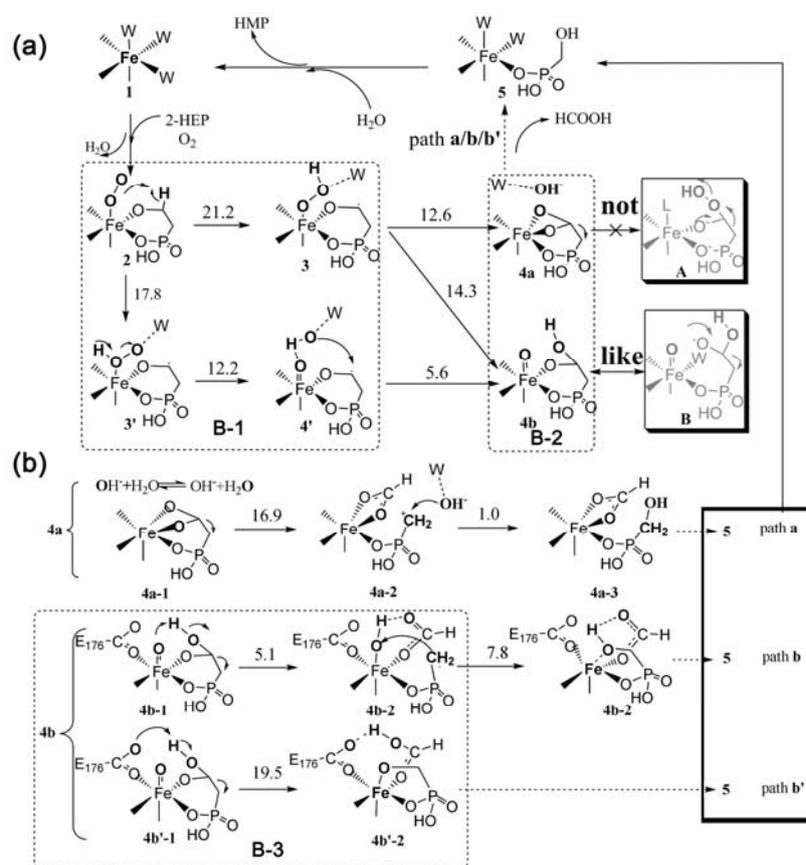


Figure S7. The total electron density differences of species **4a** and **4b** related to the carbon-carbon bond cleavage step; (a) species **4a** (b) species **4b**. (Isovalue: -0.001 for green, and 0.001 for yellow)



Scheme 1. (a) Suggested mechanism in the catalytic cycle of HEPD; (b) Possible carbon-carbon bond cleavage pathways. The number above the arrow refers to the related energy barrier (in kcal mol⁻¹). The species (**A** and **B**) with gray color in the square brackets are the putative species proposed in experiments.

- [1] R. M. Cicchillo, H. Zhang, J. A. V. Blodgett, J. T. Whitteck, G. Li, S. K. Nair, W. A. van der Donk, W. W. Metcalf, *Nature* **2009**, *459*, 871-874.
- [2] A. Sali, T. L. Blundell, *J. Mol. Biol.* **1993**, *234*, 779-815.
- [3] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph.* **1996**, *14*, 33-38.
- [4] B. Brooks, R. Brucoleri, B. Olafson, D. States, S. Swaminathan, M. Karplus, *J. Comput. Chem.* **1983**, *4*, 187-217.
- [5] D. Bashford, Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, *J. Phys. Chem. B* **1998**, *102*, 3586-3616.
- [6] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926-935.
- [7] C. L. B. III, M. Karplus, *J. Chem. Phys.* **1983**, *79*, 6312-6325.
- [8] a) C. J. Cramer, D. G. Truhlar, *Phys. Chem. Chem. Phys.* **2009**, *11*, 10757-10816; b) C. Rong, S. Lian, D. Yin, A. Zhong, R. Zhang, S. Liu, *Chem. Phys. Lett.* **2007**, *434*, 149-154; c) A. Altun, S. Shaik, W. Thiel, *J. Am. Chem. Soc.* **2007**, *129*, 8978-8987; d) A. Chanda, F. T. de Oliveira, T. J. Collins, E. Munck, E. L. Bominaar, *Inorg. Chem.* **2008**, *47*, 9372-9379.
- [9] a) K. Hemelsoet, F. De Vleeschouwer, V. Van Speybroeck, F. De Proft, P. Geerlings, M. Waroquier, *ChemPhysChem* **2011**, *12*, 1100-1108; b) H. Chen, M. Ikeda-Saito, S. Shaik, *J. Am. Chem. Soc.* **2008**, *130*, 14778-14790; c) M. Swart, A. R. Groenhof, A. W. Ehlers, K. Lammertsma, *J. Phys. Chem. A* **2004**, *108*, 5479-5483; d) M. Swart, *J. Chem. Theory Comput.* **2008**, *4*, 2057-2066.
- [10] R. Ahlrichs, M. Bär, M. Häser, H. Horn, C. Kölmel, *Chem. Phys. Lett.* **1989**, *162*, 165-169.
- [11] a) O. Christiansen, H. Koch, P. Jørgensen, *Chem. Phys. Lett.* **1995**, *243*, 409-418; b) F. Weigend, M. Häser, H. Patzelt, R. Ahlrichs, *Chem. Phys. Lett.*

- 1998**, 294, 143-152; c) C. Hättig, A. Hellweg, A. Köhn, *Phys. Chem. Chem. Phys.* **2006**, 8, 1159-1169.
- [12] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, 7, 3297-3305.
- [13] C. Hättig, F. Weigend, *J. Chem. Phys.* **2000**, 113, 5154-5161.
- [14] C. Hättig, A. Köhn, *J. Chem. Phys.* **2002**, 117, 6939-6951.
- [15] a) F. Wennmohs, F. Neese, *Chem. Phys.* **2008**, 343, 217-230; b) F. Neese, A. Hansen, F. Wennmohs, S. Grimme, *Acc. Chem. Res.* **2009**, 42, 641-648.
- [16] a) H. Hirao, D. Kumar, L. Que, S. Shaik, *J. Am. Chem. Soc.* **2006**, 128, 8590-8606; b) D. Kumar, W. Thiel, S. P. de Visser, *J. Am. Chem. Soc.* **2011**, 133, 3869-3882; c) T. Borowski, P. E. M. Siegbahn, *J. Am. Chem. Soc.* **2006**, 128, 12941-12953.
- [17] a) H. Chen, W. Lai, S. Shaik, *J. Phys. Chem. B* **2011**, 115, 1727-1742; b) M. Radoń, K. Pierloot, *J. Phys. Chem. A* **2008**, 112, 11824-11832; c) N. Nakatani, Y. Hitomi, S. Sakaki, *J. Phys. Chem. B* **2011**, 115, 4781-4789.
- [18] a) A. D. Becke, *J. Chem. Phys.* **1993**, 98, 5648-5652; b) A. D. Becke, *J. Chem. Phys.* **1993**, 98, 1372-1377; c) T. Wang, G. Brudvig, V. S. Batista, *J. Chem. Theory Comput.*, 6, 755-760.