Theoretical Studies of Photoproteins and Non-Heme Iron Enzymes: Electronic Structures and Reaction Processes

Naoki Nakatani

2010

Preface

Nowadays, biological systems and biological phenomena attract wide and considerable interests for not only in life science such as medicine and pharmacology, but also chemistry and physics. To make further developments of science and technology concerning biological systems and their functions, well understanding and correct knowledge about the mechanism of biological phenomena are indispensable, which is one of central issues in recent science.

Many biological phenomena are induced by the activities and functions of proteins, indicating that proteins are one of the most important elements in biological systems. Thus, the investigation of the protein functions is crucial to understand the many biological systems and biological phenomena at the molecular level. However, it is not easy to experimentally characterize the protein functions because protein structures and their functions are very complicated in general. In this regard, theoretical studies of proteins are indispensable to investigate the molecular mechanisms of biological phenomena.

Because of recent developments of theories and computer technologies, theoretical studies can approach larger and larger molecules including proteins. For instance, molecular dynamics (MD) simulation has succeeded to investigate the dynamical behavior of proteins which is one of the essential origins of protein functions. The MD simulation can handle more than millions of atoms, which is one of the strongpoint of this method. However, the MD simulation cannot describe the bond formation and the bond cleavage processes which are essential to many enzymatic reactions. Also, the MD simulation cannot approach electron transition and electron transfer processes which are also essential to photobiological phenomena. These biological phenomena should be investigated by electronic structure theory based on quantum chemistry. For instance, hybrid method of quantum mechanics (QM) and molecular mechanics (MM) has been often applied to theoretical investigation of

the enzymatic reactions and photobiological phenomena. In this method, the time-consuming electronic structure theory is applied to the important moiety, in which electronic process occurs, and the less time-consuming MM method is applied to the remaining moiety to incorporate steric and electrostatic interactions of the remaining moiety.

In this thesis, the author carried out theoretical studies of the photobiological phenomena and the enzymatic reactions which occur through complicated electronic process, to clarify essential roles of protein functions. Because accurate electronic structure theory must be applied to the complicated electronic structures, the modern electronic structure theories such as DFT (density functional theory), SAC/SAC-CI (symmetry-adapted cluster-configuration interaction), and CASSCF/CASPT2 (complete active space self consistent field followed by second order perturbation theory) were employed here, where the protein environmental effects were incorporated with the MM method if necessary. In chapter 1, chemiluminescence and bioluminescence of firefly luciferin are theoretically investigated with SAC/SAC-CI method to elucidate the origin of yellow-green bioluminescence. In chapter 2, the author theoretically investigates the color-tuning mechanism of firefly luciferase with the QM/MM method and proposes theoretical mutation of several amino acids to present red-colored luciferase, where the SAC/SAC-CI method is employed for the QM region and the MM method is used to consider protein electrostatic field. In chapter 3, the geometry, electronic structure, and IR frequency of the active site in iron-sulfur cluster-free hydrogenase (Hmd) are theoretically investigated with the DFT method. In chapter 4, the author describes theoretical investigation of the dioxygen binding process of iron(III) catechol dioxygenase, where the CASSCF/CASPT2 method is employed. In chapter 5, the author carries out the multi-state (MS) CASPT2 study of the ligand-to-metal charge-transfer excitation in iron(III) catechol dioxygenase and its functional models to clarify the electronic structures and their reactivity for dioxygen molecule.

The studies presented in this thesis were carried out at the Department of Synthetic Chemistry and Biological Chemistry, Graduate School of Engineering, Kyoto University from 2004 to 2007 and the Department of Molecular Engineering, Graduate School of Engineering, Kyoto University from 2007 to 2010. The author would like to express his deepest appreciation to Prof. Shigeyoshi Sakaki for his fruitful discussions and encouragement. His helpful suggestions were indispensable to carry out the studies in chapters 3, 4, and 5. The author would like to express his deepest appreciation to Prof. Hiroshi Nakatsuji for his fruitful suggestions and encouragement, too. His elegant advices are fully incorporated to studies in chapters 1 and 2. The author would like to express his sincere gratitude to Associate Prof. Dr. Hirofumi Sato and Lecturer Dr. Jun-ya Hasegawa for their kind discussion and fruitful suggestions. The author would like to thank to Assistant Prof. Dr. Yoshihide Nakao for his kind technical supports. The author would like to express sincere acknowledgement to Prof. Masayoshi Ehara for his kind discussion and helpful suggestions about SAC/SAC-CI calculations. His advices were indispensable to carry out the studies in chapters 1 and 2. The author would like to express his deepest gratitude to Prof. Takuzo Funabiki and Associate Prof. Dr. Yutaka Hitomi in Department of Molecular Chemistry and Biochemistry, Faculty of Science and Engineering, Doshisha University, for their collaboration, interesting discussion, and fruitful suggestions. Their suggestions were indispensable to carry out the studies in chapters 4 and 5.

The author is grateful to Dr. Kazuyoshi Fujimoto, Dr. Daisuke Yokogawa, and Dr. Yu-ya Ohnishi for their scientific talks on various occasions. The author also thanks all members of research groups of Prof. Shigeyoshi Sakaki and Prof. Hiroshi Nakatsuji.

The author thanks Japan Society for the Production of Science (JSPS) for financial support (Grant-in-Aid for JSPS Fellows).

Finally, the author wishes to express his sincere gratitude to his parents Chikage

Nakatani and Hitomi Nakatani for their continuous understanding, encouragement, and support from all sides.

Naoki Nakatani

February 2010

Contents

Chapter 1

General Introduction

1. Interests in Biological Systems: Protein Functionality

Proteins, one of the most important elements in biological systems, are responsible to various biological phenomena. Protein is a very big molecule which consists of more than hundreds of amino acids combined by peptide bonds, but it contains only 20 kinds of amino acids. This means that the formation of the varieties of amino acid sequence is the origin of various protein structures and functions. Enzymes which are one kind of proteins catalyze various chemical reactions in biological system with high reactivity and high selectivity under mild reaction conditions, compared to artificial homogeneous catalysts. These features are understood in terms of key-keyhole relation between enzyme (keyhole) and substrate (key).[1] This idea is also applied to pharmacology because many diseases arise from troubles of enzymatic reactions.[2] The molecule that is structurally similar to original substrate may bind with the active pocket to inhibit the enzymatic reaction. This means that such molecule is potentially useful for medicine because it might be able to control and/or suppress abnormal enzyme activity.

Of huge numbers of enzymes, metalloenzymes which involves transition metal complex in its active site are of considerable interests. They are widely found in important biological systems and play crucial roles in biological functions.[3] For instance, various oxygenases and oxidases catalyze biodegradation which occurs via chemical bond cleavage of saturated and/or aromatic hydrocarbons. Also, metalloenzymes participate in the fixation of small inorganic molecule such as H_2, N_2, and CO_2 into organic molecules. As well known,

the transition metal complex has flexible electronic structure in general because it can takes various oxidation states and spin states depending on coordinated ligands. In addition, protein environment provides characteristic reaction field available for fine tuning of reactivity. All these are responsible for metalloenzyme function.

Photobiological phenomena are good examples of protein functions. For instance, vision of eye is crucial for us, which originates from the activity of retinal protein.[4] This protein function is understood to be a biological photo-sensor. Bioluminescence is other interesting photobiological phenomenon which is also induced by protein function. Bioluminescence is also important for recent bioimaging technique.[5] Experimentalists are successfully using this technique to directly observe activity of cell. This technique plays indispensable roles to make considerable development of modern molecular biology. In the bioimaging technique, a "marker" protein, that can exhibit fluorescence or luminescence, is inserted to the base sequence of target protein at gene level and coded simultaneously with the target protein. By detecting fluorescence or luminescence from marker protein, experimentalists can understand where the target protein is in cell. Green fluorescent protein (GFP) is the most famous marker to observe the activity of protein, because of its strong green fluorescence.[6] Firefly luciferase (Luc) is used nowadays to quantitatively analyze ATP consumption in brain cell, because of its strong yellow-green bioluminescence, which is well known as the mysterious light of firefly.[7] In the biological maker, the light color tuning is of considerable important: For instance, strong red color emission is quite useful because red light is less scattered by water in cell to increase the quantum yield. In this regard, various GFP mutants that can exhibit red light were proposed,[8] and widely used in many bioimaging experiments.

Understanding the biological systems and the biological phenomena is one of the most

important and interesting subjects in science, as discussed above. Such knowledge is indispensable for further development of medicine, pharmacology, physics, and chemistry. Interestingly, reactions and behavior of molecules in biological systems are completely different from those of free molecules. Thus, not only molecules but also total biological systems including surrounding media should be taken into account in investigation of the biological systems and biological phenomena. However, it is not easy to experimentally investigate them at molecular level because biological systems are very complicated in general. In this regard, it is worthy of theoretically investigating the biological systems and biological phenomena with molecular theory.

2. Theoretical Chemistry in Biology

Classical molecular dynamics (MD) simulations have been succeeding in investigating the global motion and/or dynamics of protein,[9] which is one of the essential aspects of protein functionality. For instance, it is often employed to investigate protein folding, substrate recognition, and signaling. The classical MD simulation is able to handle a system including even millions of atoms. This is one of the strongest points for theoretical investigation of biological systems. However, it cannot describe changes of electronic structure in general. As a result, the classical MD simulation cannot be applied to the event including bond formation and bond cleavage because such events occur with concomitant reorganization of electronic structure. Also, the classical MD simulation is not useful to investigation of electron transition and electron transfer, which are essential and crucial processes in photobiological phenomena, because these events induce significant changes of electronic structure.

Considering above-mentioned issues, electronic structure theory should be employed to

theoretical studies of enzymatic reactions and photobiological phenomena. However, it is very difficult to present the electronic structure of the whole biological system because electronic structure calculation of such system needs huge computational costs. To avoid such difficulty, a small model system which involves only important active site is employed in many theoretical studies based on electronic structure theory. Such model is called active site model. Though the use of the active site model is effective to considerably reduce computational costs, serious problem that the protein environmental effect is almost neglected is in this model. To incorporate the environmental effect of protein, the QM/MM method (QM = quantum mechanics; MM = molecular mechanics) has been often employed to theoretical investigations of enzymatic reactions and photobiological phenomena. This is hybrid method consisting of quantum mechanics and molecular mechanics, in which the active site of protein is investigated by the QM method and the other region of protein surrounding active site is investigated by the MM method. This procedure successfully incorporates the environmental effect of protein with moderate computational costs because the time-consuming QM calculation is applied only to the relatively small active site and the much less time-consuming MM method is applied to very big protein environment. Actually, a large number of QM/MM calculations of biological systems have been reported recently.[10]

Theoretical investigation generally needs structural information of target system. In a small molecule, we can easily obtain the structural information from geometry optimization by theoretical method, nowadays. In a large molecule such as protein, however, it is very difficult to theoretically present reliable geometry because there are many local energy minima, as well known. Thus, the experimental structure of protein is necessary to theoretical investigations of biological systems. Recent crystallographic and NMR spectroscopy can provide us with three-dimensional structure of protein. This structural

information is quite useful to theoretical investigation of biological system with the QM/MM method. More than 60 thousand experimental structures have been reported in protein data bank, as well known.[11] In the protein data bank, two kinds of geometrical data from X-ray crystallography and NMR spectroscopy are presented. NMR spectroscopy is able to determine the protein structure including hydrogen atoms. However, this method is limited to only small protein or specific part of protein. Though many X-ray structures are presented in protein data bank, the positions of hydrogen atoms are not reported by the X-ray diffraction data. Because hydrogen atoms are very important to construct protein structure through hydrogen bonding interactions these data cannot be directly employed in theoretical study.

Though these structural data both have weak points, they are very useful to theoretical study and the combination of the QM/MM method and these experimental studies leads to successful investigation of biological systems. Because of these significant advancements in both theoretical and experimental studies, the author expects to successfully elucidate biological phenomena with theoretical methods at molecular level.

3. Quantum Mechanics for Biological Systems: To Investigate Enzymatic Reactions and Photobiological Phenomena

As mentioned above, photobiological phenomena occur with electron transition between ground and excited states. Thus, it is necessary to employ electronic structure theory which can present reliable results about ground and excited states. Though the density functional theory (DFT) presents reliable results about ground state, serious problems have been pointed out in excited state.[12] On the other hand, symmetry adapted cluster (SAC)[13] expansion method followed by configuration interaction (SAC-CI)[14] can reliably

investigate both ground and excited states. Actually, the SAC-CI method has been successfully applied to the excited states of various systems including photobiological systems.[15] SAC-CI-calculated excitation energies agree well with experimental spectra in many cases. Thus, this method is believed to be useful and effective to investigation of photobiological phenomena at molecular level.

To investigate enzyme functions, density functional theory (DFT)[16] has been often employed as the QM method in the QM/MM calculation, which successfully evaluates the changes of geometry, energy, and electronic structure by various enzymatic reactions in many cases. However, it is not perfect, as follows: Metalloenzymes involve transition metal complex in their active site, which plays crucial roles in various catalytic reaction. Interestingly, Fe is often found in biological systems. Other 3d metals (V, Cr, Mn, Co, Ni, Cu, and Zn)[3] and 4d and 5d metals (Mo and W) are also found in biological systems, but much less than Fe. Though the DFT method correctly evaluated the reaction energies of 4d and 5d transition metal complexes in general, we are afraid of some problems in the calculation of 3d transition metal complexes, when the electronic structure and wave function are very complicated due to considerably large electron correlation effects of 3d-electrons. Also, 3d metals, such as Mn, Fe, Ni, and Cu, exhibit varieties of oxidation and spin states. Such flexible electronic structure is considered to be essential to the enzymatic reaction. Thus, we must carefully carry out theoretical calculations of the enzymatic reaction including these 3d transition metal complexes. In such case, not only DFT method but also complete-active-space self-consistent-field (CASSCF)[17] theory followed by second order perturbation theory (CASPT2)[18] must be employed. However, the CASPT2 method requires large computational cost, correct understanding of electronic structure, and also high technique to present reliable results. As a result, this method has not been applied to

biological systems except for several pioneering works.[19]

4. Aims of This Thesis

Considering above-described situations in the theoretical study of the biological system, it is worthy of investigating the biological system with electronic structure theory which can successfully incorporate large electron correlation effects. The present author wishes to report here the theoretical investigations of phobiological phenomena and enzymatic reactions catalyzed by metalloenzymes with such sophisticated methods as SAC-CI and CASPT2.

In chapter 1, the yellow-green and the red color chemiluminescences from oxyluciferin were theoretically investigated with the SAC/SAC-CI method. Oxyluciferin is the chromophere of firefly bioluminescence but it also exhibits chemiluminescence, interestingly. In chemiluminescence, it has been proposed that the enol-oxyluciferin exhibits yellow-green luminescence and the keto-oxyluciferin exhibits red luminescence.[20] Because of the similarity between chemiluminescence and bioluminescence, the yellow-green bioluminescence of firefly luciferase has been believed for long time to be the emission from enol-oxyluciferin. However, recent experiment showed that the oxyluciferin keeps keto form in luciferase and the yellow-green bioluminescence occurs from the keto-oxyluciferin.[21] The alternative mechanisms which explain the yellow-green bioluminescence from the keto-oxyluciferin were proposed recently.[22] The purpose of chapter 1 is to theoretically investigate the emission energies of several oxyluciferin isomers in solution and protein atmospheres to elucidate what species participates in the yellow-green bioluminescence of firefly.

In chapter 2, the color-tuning mechanism of firefly luciferase was theoretically proposed based on the SAC/SAC-CI method. Recently, firefly bioluminescence has been

applied to quantitative analysis of ATP in brain cell.[7] In such experiment, the red colored bioluminescence is useful because the red light is less scattered by water molecule in cell, as mentioned above. There are about two thousand kinds of firefly, and the color of the emission varies from green (~530 nm) to red (~635 nm).[23] Thus, the color-tuning mechanism of firefly luciferase is of considerable interest. In this chapter, the author analyzed the electrostatic interactions between the firefly oxyluciferin and the surrounding proteins to clarify what amino acids play key roles to control the emission energy of oxyluciferin. Furthermore, based on this knowledge, the author proposed the theoretical mutant of firefly luciferase which exhibits red colored bioluminescence.

In chapter 3, the active center of iron-sulfur-cluster-free hydrogenase (Hmd)[24] was theoretically investigated, where the DFT method was employed because the ground state properties such as geometry and vibration frequency were mainly discussed. Hydrogenase catalyzes reversible H_2 activation. Interestingly, Hmd contains mononuclear low-spin iron center,[25] which is considerably different from the active center of typical [NiFe]-hydrogenase including binuclear iron-nickel complex. However, there are many issues to be investigated in the Hmd active center, as follows: (1) It is not clear which of iron(0) and iron(II) oxidation states the iron center takes. (2) One ligand is still unknown. And, (3) it is still unclear with which of monodentate pyridinol form and bidentate acylpyridinol the pyridinol cofactor coordinates with the iron center. The purpose of chapter 3 is to theoretically investigate the geometry, iron oxidation state, and electronic structure of Hmd active site.

In chapter 4, dioxygen binding process of non-heme iron(III) center in intra-diol catechol dioxygenase was theoretically investigated. The author employed here the CASSCF/CASPT2 method to incorporate multi-configurational character of the Fe-O_2 interaction. In this process, two alternative mechanisms were experimentally proposed; one

is called “Oxygen Activation”[26] and the other is called “Substrate Activation”.[27] In the oxygen activation mechanism, the dioxygen molecule first attacks the iron(III) center. In the substrate activation mechanism, on the other hand, the charge transfer occurs from the catecholate moiety to the iron(III) center to afford the iron(II)—semiquinonate species, in which the radical center of semiquinonate reacts with dioxygen molecule to form the C-O bond first. However, because of strong electron correlation effects based on the interaction between iron(III) center and dioxygen molecule, it has not been theoretically discussed well the dioxygen binding process. The purpose of this chapter is to elucidate which mechanism is correct.

In chapter 5, the author theoretically investigated the ligand-to-metal charge-transfer (LMCT) excitation of the native iron(III) catechol dioxygenase and its functional model complexes. Because it is well accepted that the high-spin iron(III) center does not react with dioxygen molecule in general, the dioxygen activation by the iron(III) catechol dioxygenase is of considerable interest. Based on the idea that the charge transfer occurs from the catecholate moiety to the iron(III) center to afford the iron(II)—semiquinonate species which is expected to be reactive for dioxygen molecule, it has been believed in the experimental field that the reactivity for dioxygen molecule strongly depends on the contribution of iron(II)—semiquinonate character to the ground state.[27b] Actually, iron(III) catechol dioxygenase and its functional models exhibit the LMCT absorption in the range from near-infrared to visible region and its excitation energy clearly relates to the reactivity for dioxygen molecule.[28] Thus, it is worthy of theoretically investigating how much the LMCT excited state contribute to the ground state. However, the active site of the iron(III) catechol dioxygenase possesses extremely complicated electronic structure based on considerably large non-dynamical electron correlation effects and quasi-degeneracy problem. The purposes of

chapter 5 are to theoretically investigate both ground and excited states of the iron(III) catechol dioxygenase and its functional models with the CASSCF/CASPT2 method which incorporates dynamical and non-dynamical electron correlation effects and quasi-degeneracy problem, to discuss the relationship among the LMCT excitation energy, the electronic structures, and the reactivity for dioxygen molecule. Such knowledge is indispensable to understand the dioxygen activation mechanism of iron(III) catechol dioxygenase and to design a new model of oxygenation enzyme.

Through these studies, the author wishes to present deep understanding of electronic processes and molecular mechanisms of protein functions in which the electronic structure plays key roles. The author also wishes to suggest that the accurate electronic structure theory is required to correctly understand the electronic processes in these protein functions.

Reference

(1) Fischer, E. *Ber. Dt. Chem. Ges.* **1894**, *27*, 2985-2993.

(2) "*Genes and Disease*" in on-line textbook of National Center of Biotechnology Information (NCBI).

(3) Ochiai, E. *Bioinorganic Chemistry: A Survey*; Elsevier Academic Press, Burlington, 2008.

(4) Mathies, R. A.; Lugtenburg, J. *Handbook of Biological Physics*; Stavenga, D.G., Grip, W. J. D., Pugh, E. N. Eds.; Elsevier Science B. V., Amsterdam, 2000.

(5) (a) Haraguchi, T. *Cell Struct. Funct.* **2002**, *27*, 333. (b) Contag, C. H.; Bachmann, M. H. *Annu. Rev. Biomed. Eng.* **2002**, *4*, 235.

(6) Tsien, R. Y. *Annu. Rev. Biochem.* **1998**, *67*, 509-544.

(7) (a) Campbell, A. K. *Chemiluminescence: Principles and Applications in Biology and Medicine*; VHC, Chichester, 1988. (b) Naylor, L. H. *Biochem. Pharm.* **1999**, *58*, 749.

(8) (a) Chudakov, D. M.; Belousov, V. V.; Zaraisky, A. G.; Novoselov, V. V.; Staroverov, D. B.; Zorov, D. B.; Lukyanov, S.; Lukyanov, K. A. *Nat. Biotechnol.* **2003**, *21*, 191. (b) Marchant, J. S.; Stutzmann, G. E.; Leissring, M. A.; LaFerla, F. M.; Parker, I. *Nat. Biotechnol.* **2001**, *19*, 645. (c) Ando, R.; Hama, H.; Yamamoto-Hino, M.; Muzuno, H.; Miyawaki, A. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12651.

(9) *Computational Structural Biology: Methods and Applications*; Schwede, T., Peitsch, M. C. Eds., World Scientific, Singapore, 2008.

(10) Senn, H. M.; Thiel, W. "QM/MM Methods for Biological Systems" in Atomistic Approaches in Modern Biology (Topics in Current Chemistry); Reiher, M. Ed., Springer Berlin Heidelberg, New York, 2007.

(11) RCSB *Protein Data Bank* http://www.rcsb.org/

(12) (a) Casida, M. E.; Jamorski, C.; Casida, K. E.; Salahub, D. R. *J. Chem. Phys.* **1998**, *108*, 4439. Casida, M. E.; Salahub, D. R. *J. Chem. Phys.* **2000**, *113*, 8918. (b) Appel, F.; Gross, E. K. U.; Burke, K. *Phys. Rev. Lett.* **2003**, *90*, 043005. (c) Hsu, C.-P.; Hirata, S.; Head-Gordon, M. *J. Phys. Chem. A* **2001**, *105*, 451-458. (d) Dreuw, A.; Weisman, J. L.; Head-Gordon, M. *J. Chem. Phys.* **2003**, *119*, 2943-2946. (e) Milet, A.; Korona, T.; Moszynski, R.; Kochanski, E. *J. Chem. Phys.* **1999**, *111*, 7727. (f) Fujimoto, K.; Hayashi, S.; Hasegawa, J.; Nakatsuji, H. *J. Chem. Theory Comput.* **2007**, *3*, 605-618.

(13) Nakatsuji, H.; Hirao, K. *J. Chem. Phys.* **1978**, *68*, 2053.

(14) (a) Nakatsuji, H. *Chem. Phys. Lett.* **1978**, *59*, 362. (b) Nakatsuji, H. *Chem. Phys. Lett.* **1979**, *67*, 329. (c) Nakatsuji, H. *Chem. Phys. Lett.* **1979**, *67*, 334.

(15) Nakatsuji, H. *Computational Chemistry, Reviews of Current Trends*; Leszczynski, J., Ed.; World Scientific: Singapore, 1996; Vol. 2, pp 62.

(16) (a) Kohn, W.; Sham, L. *J. Phys. Rev.* **1965**, *140*, A 1133. (b) Parr, R. G.; Yang, W.

Density-Functional Theory of Atoms and Molecules; Oxford: New York, 1989. (c) Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory 2nd Ed.*; Wiley: Weinheim, 2001.

(17) (a) Roos, B. O.; Taylor, P. R. *Chem. Phys.* **1980**, *48*, 157-173. (b) Roos, B. O. In *Ab Initio Methods in Quantum Chemistry II*; Lawley, K. P., Ed.; Wiley: New York, 1987; p.399.

(18) (a) Andersson, K.; Malmqvist, P.-Å.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. *J. Phys. Chem.* **1990**, *94*, 5483-5488. (b) Andersson, K.; Malmqvist, P.-Å.; Roos, B. O. *J. Chem. Phys.* **1992**, *96*,1218-1226.

(19) Pierloot, K.; Kerpel, O. A. D.; Ryde, U.; Roos, B. O. *J. Am. Chem. Soc.* **1997**, *119*, 218-226.

(20) White, E. H.; Rapaport, E.; Hopkins, T. A.; Seliger, H. H. *J. Am. Chem. Soc.* **1969**, *91*, 2178.

(21) Branchini, B. R.; Murtiashaw, M. H.; Magrar, R. A.; Portier, N. C.; Ruggiero, M. C.; Stroh, J. G. *J. Am. Chem. Soc.* **2002**, *124*, 2112.

(22) (a) Branchini, B. R.; Southworth, T. L.; Murtiashaw, M. H.; Magyer, R. A.; Gonzalez, S. A.; Ruggiero, M. C.; Stroh, J. G. *Biochemistry* **2004**, *43*, 7255. (b) McCapra, F.; Gilfoyle, D. J.; Young, D. W.; Church, N. J.; Spencer, P. *Bioluminescence and Chemiluminescence. Fundmental and Applied Aspects*; Campbell, A. K., Kricka, L. J., Stanley, P. E. Eds.; John Wiley & Sons: Chichester, U.K., 1994; p.387. (c) Orlova, G.; Goddard, J. D.; Brovko, L. Y. *J. Am. Chem. Soc.* **2003**, *125*, 6962.

(23) (a) Wood, K. V. *Photochem. Photobiol.* **1995**, *62*, 662. (b) Wood, K. V.; Lam, Y. A.; Seliger, H. H.; McElroy, W. D. *Science* **1989**, *244*, 700. (c) Viviani, V. R.; Bechara, E. J. H. *Photochem. Photobiol.* **1995**, *62*, 490. (d) Viviani, V. R. *Cell. Mol. Life Sci.* **2002**, *59*,

1833. (e) Ugarova, N. N. *Biolum. Chemilum.* **1989**, *4*, 406.

(24) (a) Zirngibl, C.; Hedderich, R.; Thauer, R. K. *FEBS Lett.* **1990**, *261*, 112. (b) Schwörer, B.; Thauer, R. K. *Arch. Microbiol.* **1991**, *155*, 459. (c) Ma, K.; Zirngibl, C.; Linder, D.; Stetter, K. O.; Thauer, R. K. *Arch. Microbiol.* **1991**, *156*, 43.

(25) Shima, S.; Lyon, E. J.; Thauer, R. K.; Mienert, B.; Bill, E. *J. Am. Chem. Soc.* **2005**, *127*, 10430.

(26) (a) Funabiki, T.; Fukui, A.; Hitomi, Y.; Higuchi, M.; Yamamoto, T.; Tanaka, T.; Tani, F.; Naruta, Y. *J. Inorg. Biochem.* **2002**, *91*, 151-158. (b) Hitomi, Y.; Yoshida, M.; Higuchi, M.; Minami, H.; Tanaka, T.; Funabiki, T. *J. Inorg. Biochem.* **2005**, *99*, 755-763. (c) Higuchi, M.; Hitomi, Y.; Minami, H.; Tanaka, T.; Funabiki, T. *Inorg. Chem.* **2005**, *44*, 8810-8821.

(27) (a) Que, L., Jr.; Lipscomb, J. D.; Muenck, E.; Wood, J. M. *Biochem. Biophys. Acta.* **1977**, *485*, 60-74. (b) Que, L., Jr.; Kolanczyk, R. C.; White, L. S. *J. Am. Chem. Soc.* **1987**, *109*, 5373-5380. (c) Cox, D. D.; Que, L., Jr. *J. Am. Chem. Soc.* **1988**, *110*, 8085-8092.

(28) Hitomi, Y.; Yoshida, M.; Higuchi, M.; Minami, H.; Tanaka, T.; Funabiki, T. *J. Inorg. Biochem.* **2005**, *99*, 755-763.

Chapter 2

Red Light in Chemiluminescence and Yellow-Green Light in Bioluminescence: Color-Tuning Mechanism of Firefly, *Photinus pyralis*, Studied by the Symmetry-Adapted Cluster-Configuration Interaction Method

2.1 Introduction

Bioluminescence is a well-known photo-biological phenomenon in which chemical energy is converted into photon energy. Firefly luminescence is a typical example of bioluminescence. There are about two thousand types of firefly worldwide, and the color of the emission varies from green (~530 nm) to red (~635 nm) depending on the species.[1] The enzyme firefly luciferase (Luc) is also used in molecular imaging as a reporter of gene expression.[2] Therefore, the underlying molecular mechanism of color-tuning must be clarified.

The mechanism of firefly bioluminescence proposed in previous studies[3,4] is summarized in Figure 1a. Luc binds firefly luciferin ((S)-2-(6-hydroxy-2-benzothiazolyl)-2-thiazoline-4-carboxylic acid, D-LH_2) in its active site and catalyzes the formation of luciferyl adenosine monophosphate (AMP). Initiated by the attachment of O_2 to the thiazoline ring, a dioxyetanone intermediate is formed via intramolecular chemically initiated electron-exchange luminescence (CIEEL).[5] In the final step, detachment of CO_2 generates oxyluciferin ($OxyLH_2$) in an electronically excited state, which produces a yellow-green emission in the case of the North American firefly (*Photinus pyralis*).[3,6] $OxyLH_2$ also exhibits chemiluminescence, with a mechanism very similar to that of bioluminescence, as

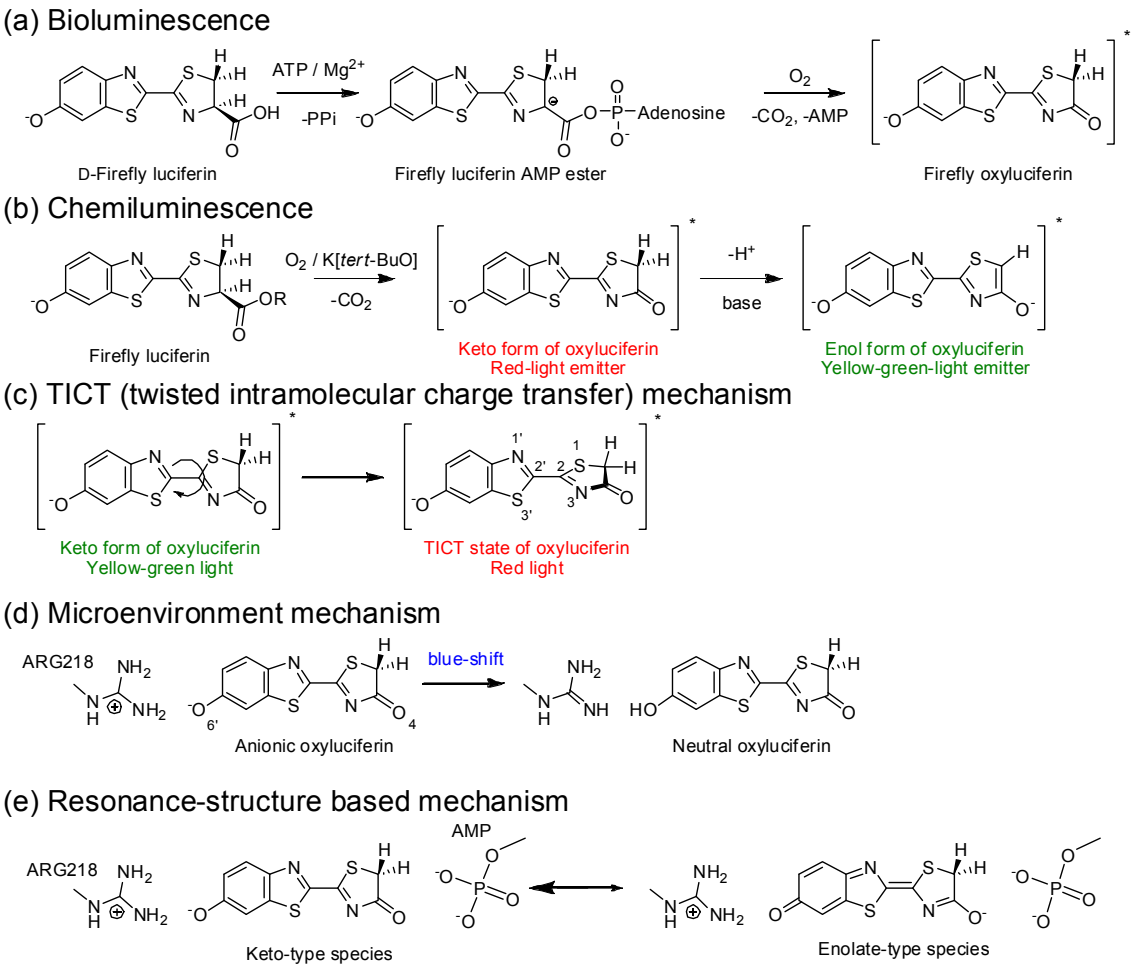


Figure 1. Proposed mechanism for (a) bioluminescence and (b) chemiluminescence of the firefly.[7] Other mechanisms for firefly bioluminescence: (c) TICT (twisted-intramolecular charge-transfer) mechanism,[11] (d) microenvironmental mechanism,[12-14] and (e) resonance structure-based mechanism (see text for details).[10]

shown in Figure 1b. It has been established that $OxyLH_2$ is in the keto form in dimethyl sulfoxide (DMSO) solution containing a small amount of potassium *tert*-butoxide (K[*tert*-BuO]). It is transformed into the enol form in DMSO solution with a large amount of K[*tert*-BuO].[6-8] The keto and enol isomers emit red (~620 nm, 1.97 eV) and green (~560 nm, 2.20 eV) light, respectively.[7] Because of the similarity between chemiluminescence and

bioluminescence, the mechanism of color tuning in bioluminescence was believed to be the same as that in chemiluminescence.[6,7]

Recently, Branchini and co-workers[9] reported that 5,5-dimethyloxyluciferin that was constrained to exist only in the keto form produced green emission inside Luc.[9,10] This indicates that the color of the luminescence can be controlled only in the keto form. It is not necessary to invoke keto-enol tautomerism to explain the color of the bioluminescence.

At present, there are three alternative explanations. The first is the twisted intramolecular charge-transfer (TICT) state mechanism proposed by McCapra and co-workers and shown in Figure 1c.[11] Their semi-empirical AM1 calculation suggests that twisting the C2-C2' angle stabilizes the emitting state and produces the red emission. The planer structure that produces the yellow-green emission is a saddle point on the potential surface. The second explanation is the 'microenvironment mechanism', which is shown schematically in Figure 1d.[12-14] In the keto form, the protonation state of the O6' atom and Arg 218 (polarity of the hydrogen bond) is especially important.[14] The third proposal is related to the resonance structure of the excited $OxyLH_2$ in keto and enolate forms, as shown in Figure 1e.[10] The weights of the two resonance structures are equivalent in chemiluminescence, thereby stabilizing the emitting state. In the Luc environment, however, the local interaction constrains the resonance structure more to the left, which causes a blue-shift of the emission.[10] It is now necessary to characterize the electronic and molecular structures of the emitting state in the Luc environment and to clarify the mechanism of color tuning.

Since luminescence is a transient phenomenon during the photochemical process, reliable quantum mechanical (QM) calculations would provide useful information on the nature of the excited state involved. Based on the X-ray structure[15] and working model provided by the previous experimental studies,[16,17] we performed molecular mechanical

(MM) and QM calculations to determine the structure of the $OxyLH_2$-Luc binding complex in the excited state. For studying luminescence originating from molecular excited state, electron-correlation effects are essential for reproducing the excitation/emission energy in QM calculations. We used the Symmetry-Adapted Cluster (SAC)[18]/SAC-Configuration Interaction (CI)[19] method[20] for balanced description of the electron correlation effects in both ground and excited states. One of the computationally inexpensive alternative approaches may be time-dependent density-functional theoretical (TDDFT) calculation. However, the accuracy of the TDDFT is out of the required accuracy for photobiology, so that the TDDFT is useless for clarifying the underlying principle behind the photo-biological phenomena. There are already several reports concerning the difficulties in the TDDFT method.[21] In our experience, TD-B3LYP calculations failed to systematically reproduce the excitation energies of retinal proteins.[21e] Since the excitation phenomena of $OxyLH_2$ in the Luc environment are such an unresolved problem, we did not perform TDDFT calculations to avoid introducing unnecessary confusions in the analyses.

In the present paper, we study with the SAC/SAC-CI method the chemiluminescence in DMSO solution and the bioluminescence in Luc. We will show that the keto form of $OxyLH_2$ produces red light in chemiluminescence and yellow-green in bioluminescence. We also examine the TICT state, the possibility of the enol form, and the effect of the local environmental on the protonation state~ resonance structure of $OxyLH_2$.

2.2 Computational details

2.2.1 Quantum mechanical calculations

In the study of chemiluminescence, we examined several structural isomers and tautomers of $OxyLH_2$ in different protonation states as shown in Figure 2. We adopted a

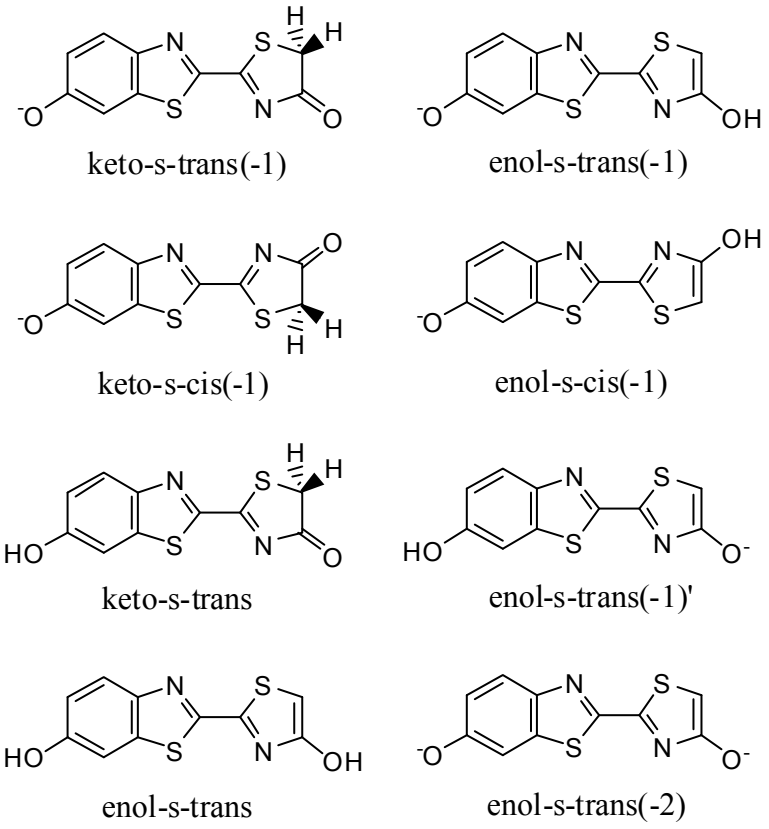


Figure 2. Structural isomers and tautomers of $OxyLH_2$ in different protonation states investigated in the present study.

conventional nomenclature[14] in which "keto" and "enol" indicate keto and enol (enolate) forms in the thiazol ring, respectively. Number in the parenthesis shows total charge of $OxyLH_2$. For enol-s-trans(-1)', the prime indicates a protonation state different from that of enol-s-trans(-1). The geometries of their first excited states were optimized using the CI-Single (CIS) method and single-point SAC-CI calculations were performed at the optimized geometries for obtaining the fluorescence energies. Since the wave functions of the first excited states were well described by the one-electron transitions from HOMOs to LUMOs, the CIS geometries were qualitatively correct. Actually, we performed also the SAC-CI geometry optimizations for the first excited states of keto-s-trans(-1) and enol-s-trans(-1) in a gas phase. In Figure 3 and Table A1 (Appendix), bond lengths obtained

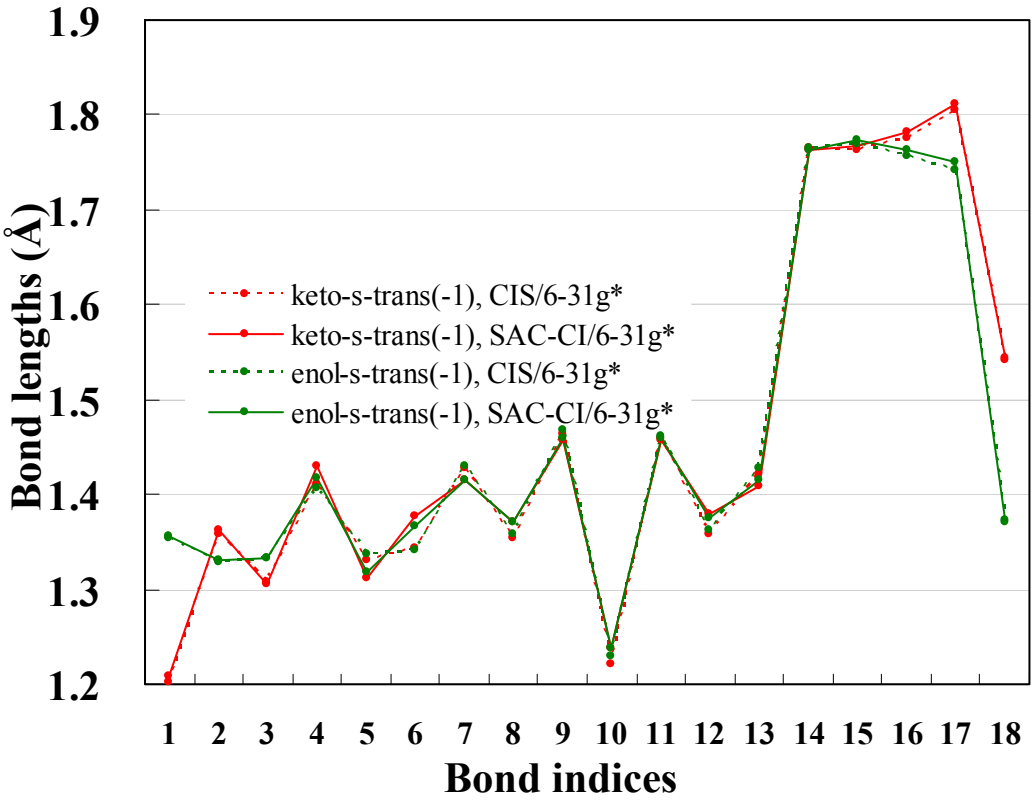


Figure 3. Bond lengths of keto and enol $OxyLH_2$ calculated by CIS and SAC-CI method in the gas phase.

by CIS and SAC-CI methods are summarized. The RMS deviation between the CIS and SAC-CI results was 0.014 and 0.010 Å in the keto and enol forms, respectively. The largest deviation is in the R6 (N-C) bond in both keto and enol forms, and the CIS bond length is about 0.03 Å shorter than the SAC-CI one. With the CIS and SAC-CI geometries, the calculated SAC-CI fluorescence energies of the keto form were 1.97 and 1.84 eV, respectively, and for the enol form, they were 2.04 and 2.01 eV, respectively. The correlation effect on the geometry of the excited state was very small (0.03 eV) in the enol form, but it was not negligible in the keto form (0.13 eV = 3.07 kcal/mol). However, such differences were fortunately small enough not to affect the conclusions of the present paper. In the SAC-CI calculations, we performed perturbation selection for the double excitation

operators[22] and the energy thresholds of 5.0×10^{-6} and 5.0×10^{-7} hartree (Level Two) were used for the ground and excited states, respectively. These calculations were performed in both the gas phase and DMSO solution. To mimic the DMSO environment, a polarized continuum model (PCM)[23] was used to take into account the solvation effects. The dielectric constant of DMSO (ε=46.7) was applied. The experimental solution also included K[*tert*-BuO]: 10^{-5} M LH_2 and 0.005–0.5 M K[*tert*-BuO] in DMSO solution.[6,7] To examine the effect of the coordination to the counter ion, the K^+ ions were explicitly included in the QM models of the charged species: keto-s-trans(-1), enol-s-trans(-1), enol-s-trans(-1)', and enol-s-trans(-2). The geometry of the $OxyLH_2$-K^+ complex was also optimized using the CIS method with the PCM. For the basis sets, we used Huzinaga–Dunning double-zeta basis sets[24] augmented with the single-polarization d-function,[24] D95(d), for the calculations of both geometries and fluorescence energies.

To evaluate the possibility of a TICT excited state, we studied the potential energy curve for twisting of the C2-C2' bond (Figure 1c) in a gas phase. The geometry optimizations were done at the CIS/D95(d) level and the single-point SAC-CI calculations with D95(d) basis sets were done at the optimized geometries. The dihedral angle was rotated from 0.0° (trans) to 180.0° (cis) in steps of 10.0°. Since we used perturbation selection for the double excitation operators, we took the group sum of the excitation operators selected at each point of the potential curve (GSUM method[25]). The GSUM method guarantees a smooth potential curve despite of the selections of the excitation operators. To ensure the GSUM method accurate, we used the Pipek–Mezey localized molecular orbitals.[26]

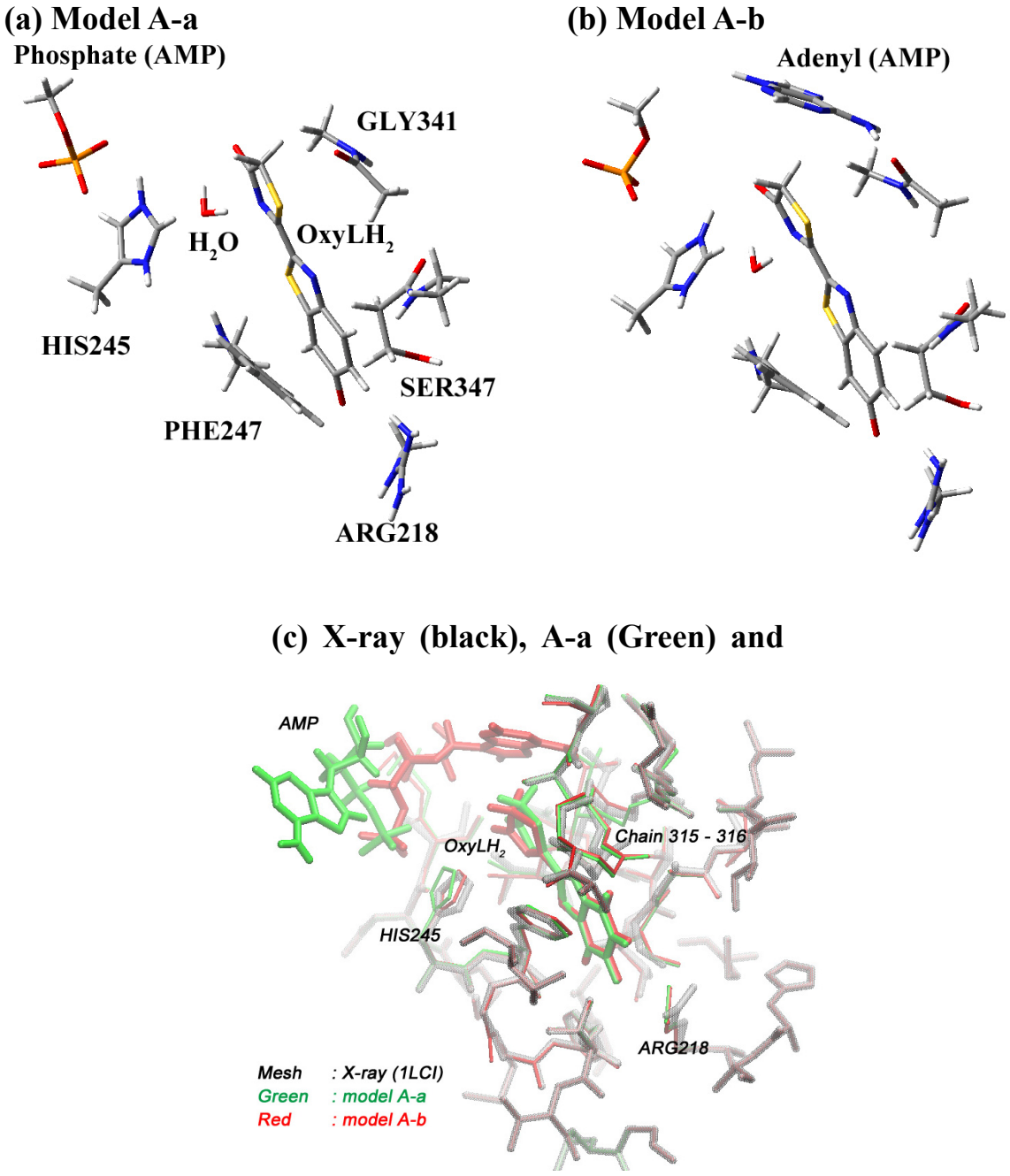


Figure 4. (a, b) Computational models treated by ab initio CIS optimization. Atoms not shown in the figures were treated by the point-charge model. (c) Superimposed structures for comparison of X-ray crystallographic structure (black, mesh), models A-a (green), and A-b (red)

2.2.2 Structural optimization of Luc-OxyLH_2 complex

Based on the X-ray structure (1LCI[15]) and the working model proposed in the experimental studies,[10,16] model A and model B (see Figure A1 in the Appendix) were constructed as initial guesses, and molecular dynamics (MD) calculations were performed to relax the structures. Both models include hydrogen-bonding with Arg218 as suggested in a previous study.[17] In model A, OxyLH_2 is placed in the pocket, so that the carbonyl oxygen in its thiazolinone ring can interact with His245. Model B has the opposite orientation. The results of MD simulations showed that Model A is much more stable than Model B, because of the hydrogen bonding between OxyLH_2 and the histidine.

The MD simulations were done with the use of the AMBER94 force-field.[27] The force-field parameters were constructed for the first excited state of OxyLH_2. The equilibrium bond lengths, bond angles, dihedral angles, and electrostatic potential charges were obtained from the geometry optimized by CIS/6-31g*. For the force constants and van der Waals parameters, we used those for a similar "atom type" defined in the AMBER94 force-field.[27] Before the MD simulations, the hydrogen atoms and some missing residues were added to the X-ray (1LCI) structure, and preliminary geometry optimizations were done for these atoms. Classical MD simulations were conducted for OxyLH_2, AMP, and selected amino acids and water molecules. If all of the atoms in the amino acids and water molecules were located more than 7.0 Å away from the luciferin, the coordinates were frozen in the simulation. To find the stable structures, the temperature of the MD simulation was raised stepwise up to 300 K. In each step, the temperature was raised by 50 K, and a 10 ps MD calculation was performed to stabilize the structures. At 300 K, a 30 ps MD calculation was made to obtain the reference structure.

After the MD simulation, we extracted two stable structures, **A-a** and **A-b,** from the

trajectory obtained with Model A. These structures were further refined by the MM and *ab initio* CIS geometry optimizations. In the CIS calculations, $OxyLH_2$, AMP, ARG218, HIS245, PHE247, GLY341, LEU342, SER347, ALA348, and the crystal water HOH45 were treated with the QM method and the rest of the protein by the classical point-charge model defined by the MM parameters. The basis sets of 6-31g* were used for $OxyLH_2$, and the 6-31g set for the rest of the QM atoms. The optimized structures are shown in Figure 4a,b, and their superposition is shown in Figure 4c. These two structures are similar to each other except for the position of the adenyl group. The phosphate group is close to His245, because of the interaction between the negative charge in the phosphate and the positive charge in the His245. Although the gas-phase structure of $OxyLH_2$ is co-planar, the C2-C2' bond is slightly twisted. For Models A-a, and A-b, the dihedral angles, N=C-C=N, are 170° and 160°, respectively.

We also compared the present optimized structure with the X-ray crystallographic structure of Luc.[15] This structure determination is possible only with the theoretical method because it is for the excited state of luciferin. As seen in Figure 4c, the residues directly interacting to $OxyLH_2$ show relatively large differences between the X-ray structure (black) and the present optimized geometry (green and black). They rearrange their conformations to enlarge the interaction with $OxyLH_2$. For example, positively charged Arg218 twists its guanidine group toward the carboxy group of $OxyLH_2$. The position of His245 depends to that of AMP. The changes in the main chain of Gly315 and 316 are due to the insertion of $OxyLH_2$ into the binding site. The cartesian coordinates of the models A-a and A-b were given in the Appendix.

The QM and MM calculations were performed using the *Gaussian* 03[28] and TINKER[29] programs, respectively.

2.3 Results and Discussion

2.3.1 Chemiluminescence: fluorescence energies of $OxyLH_2$ in keto and enol forms

The emission energies for the eight isomers calculated by the SAC-CI method are shown in Table 1. All of the isomers have a co-planar structure in the excited state, which agrees with the previous report.[14] The first excited state is characterized as a one-electron transition from HOMO (π) to LUMO (π^*) in all of the isomers considered in this study.

Firstly, we could exclude the neutral forms, keto-s-trans and enol-s-trans, from the candidates for the chemiluminescence emitter. The calculated fluorescence energies for keto-s-trans and enol-s-trans in the DMSO environment are 2.95 and 2.83 eV, respectively. These are much higher than the experimental values obtained for the keto (1.97 eV) and enol (2.20 eV) forms, respectively. Among the trans and cis isomers, we focused on the trans isomers. Although the calculated fluorescence energies for the trans and cis isomers are very close to each other, the total energy of the trans isomer in the excited state is lower than that of the cis isomer. For instance, in the DMSO environment, the keto-s-trans(-1) and enol-s-trans(-1) forms are lower than their cis isomers by 6.4 and 5.8 kcal/mol, respectively.

Regarding the keto form, the calculated emission energy for keto-s-trans(-1) was 2.10 eV with K^+ ion, which reasonably agrees with the experimental value of 1.97 eV. Thus, keto-s-trans(-1) was confirmed as the red emitter in the chemiluminescence.

For the yellow-green emitter under strongly basic conditions, the calculated emission energies of the three candidates, enol-s-trans(-1), enol-s-trans(-1)', and enol-s-trans(-2), were 2.31, 2.20, and 2.17 eV, respectively. Since all of these values are close to the experimental emission energy of 2.20 eV, we next examined the relative stability of these enol forms in the excited states. The total energy is defined as the sum of those of the potassium-$OxyLH_2$ complex and *tert*-BuO, as shown in Table 2. For instance, the total energy of the

Table 1. Chemiluminescence (fluorescence) of $OxyLH_2$ calculated by the SAC-CI method with D95(d) basis sets in the gas phase and in dimethyl sulfoxide (DMSO) solution[a]

Molecule	Environment		Emission Energy / eV	
	Optimization	Single point	SAC-CI	Exptl.
	keto forms			
keto-s-trans	Gas	Gas	3.26	
	Gas	DMSO	2.95	
keto-s-cis(-1)	Gas	Gas	1.99	
	Gas	DMSO	2.08	1.97[b]
keto-s-trans(-1)	Gas	Gas	1.96	
	Gas	DMSO	2.08	
	DMSO (with K^+)	DMSO (with K^+)	2.10	
	enol forms			
enol-s-trans	Gas	Gas	2.92	
	Gas	DMSO	2.83	
enol-s-cis(-1)	Gas	Gas	2.04	
	Gas	DMSO	2.30	
enol-s-trans(-1)	Gas	Gas	2.02	
	Gas	DMSO	2.25	
	DMSO (with K^+)	DMSO (with K^+)	2.31	2.20[c]
enol-s-trans(-1)'	Gas	Gas	—	
	Gas	DMSO	2.14	
	DMSO (with K^+)	DMSO (with K^+)	2.20	
enol-s-trans(-2)	Gas	Gas	1.85	
	Gas	DMSO	2.07	
	DMSO (with $2K^+$)	DMSO (with $2K^+$)	2.17	

[a] The term "with K^+" indicates that the K^+ ion is explicitly included in the QM calculation. [b] Experimental condition is 0.005 M potassium *tert*-butoxide (K[*tert*-BuO]) in DMSO.[7] [c] Experimental condition is 0.5 M potassium *tert*-butoxide (K[*tert*-BuO]) in DMSO.[7]

Table 2. Relative energy (kcal/mol) for the excited state of potassium-$OxyLH_2$ complex

Computational model[a]	Relative energy (kcal/mol)[b]
(a) Tight ion-pair model (without explicit DMSO)[c]	
keto form	
K[keto-s-trans(-1)] + K[*tert* -BuO]	0.0 (10.5)
enol forms	
K[enol-s-trans(-1)] + K[*tert* -BuO]	17.2 (27.7)
K[enol-s-trans(-1)'] + K[*tert* -BuO]	18.8 (29.3)
K_2[enol-s-trans(-2)] + *tert* -BuOH	-10.5 (0.0)
(b) Solvent separated ion-pair model (with explicit DMSO)[c]	
keto form	
K[DMSO][keto-s-trans(-1)] + K[*tert* -BuO] + DMSO	0.0 (6.5)
enol forms	
K[DMSO][enol-s-trans(-1)] + K[*tert* -BuO] + DMSO	17.6 (24.1)
K[DMSO][enol-s-trans(-1)'] + K[*tert* -BuO] + DMSO	20.1 (26.6)
K_2[DMSO]$_2$[enol-s-trans(-2)] + *tert* -BuOH	-6.5 (0.0)

[a] Computational model for quantum-mechanical calculation. The rest of the solvation effect was treated by the PCM(DMSO). CIS/6-31g* plus PCM calculations were performed for the excited states of the potassium-$OxyLH_2$ complex and potassium-DMSO-$OxyLH_2$ complexes, and RHF/6-31g* plus PCM(DMSO) calculations for the ground state of the *tert*-BuO and DMSO, respectively. [b] keto-s-trans(-1) is taken as the reference. The number in parentheses is the relative energy, with enol-s-trans(-2) as the reference. [c] The K^+ ion was directly coordinated by $OxyLH_2$ in "Tight ion-pair model", while "Solvent separated ion-pair model" includes a DMSO molecule between $OxyLH_2$ and K^+. See optimized structures in Figure S5.

enol-s-trans(-1) system is the sum of K[enol-s-trans(-1)] in the excited state and K[*tert*-BuO] in the ground state. RHF/6-31g* and CIS/6-31g* plus PCM calculations were carried out for optimizing the structures in the ground and excited states, respectively.

The geometry of *tert*-BuO in the ground state was optimized at RHF/6-31G* plus PCM. As Table 2 shows, enol-s-trans(-2) is most stable of the three candidates by 6.7–8.3 kcal/mol. The present calculations thus suggest that enol-s-trans(-2) is the most likely candidate for the yellow-green chemiluminescence emitter.

The relative energies shown in Table 2 explain the experimental observation that the keto to enol transformation occurred in the excited state when an excess of K[*tert*-BuO] was present.[7] The energy of the enol-s-trans(-2) system is 10.5 kcal/mol lower than that of the keto-s-trans(-1) system. Thus, excess K[*tert*-BuO] can promote the formation of the enol-s-trans(-2) form before red light emission occurs from the keto-s-trans(-1) form. In contrast, the energies of the enol-s-trans(-1) and enol-s-trans(-1)' systems are 17.2 and 18.8 kcal/mol higher than that of the keto-s-trans(-1) system.

We also examined another solvation model in which an explicit DMSO molecule exists between $OxyLH_2$ and K^+. The geometries were optimized at CIS/D95(d) plus PCM(DMSO) level (see structures in Figure A3). The DMSO molecule causes slight red shift in emission energies: 0.03, 0.04, 0.05, and 0.04 eV in keto-s-trans(-1), enol-s-trans(-1), enol-s-trans(-1)',and enol-s-trans(-2), respectively (see CIS emission energies in Table A2).

2.3.2 Bioluminescence: fluorescence energy of the keto-$OxyLH_2$ inside Luc

Firstly, we roughly determined the important residues for the color-tuning of the fluorescence. The transition to the first excited state changes the charge distribution of $OxyLH_2$, which affects the interaction energy between $OxyLH_2$ and the surrounding

molecules. Figure 5 illustrates the changes in the Mulliken atomic charges of $OxyLH_2$ (keto-s-trans(-1) form) upon the excitation, $Q_A^{Ex} - Q_A^G$. In the first excited state, the O4 atom becomes more negative, while the negative charge of the O6' is decreased upon excitation. The O4 atom is located within 5 Å of His245 (+1 charge) and AMP (–2 charge in the phosphate group), and the O6' atom has a hydrogen bond with Arg218 (+1 charge). Thus, the difference in the interaction energy between the ground and excited states (ΔE) would contribute to the shift in the emission energy caused by the Luc environment.

$$\begin{aligned}\Delta E &= \sum_{M}^{\text{Residues}} \Delta E_M \\ &= \sum_{M}^{\text{Residues}} \left\{ E_M^{Ex} - E_M^G \right\} \\ &= \sum_{M}^{\text{Residues}} \left\{ \sum_{A}^{\text{OxyLH}_2} \sum_{B \in M}^{\text{Others}} \frac{Q_A^{Ex} Q_B}{r_{AB}} - \sum_{A}^{\text{OxyLH}_2} \sum_{B \in M}^{\text{Others}} \frac{Q_A^{G} Q_B}{r_{AB}} \right\}\end{aligned} \qquad (1)$$

ΔE can be easily decomposed into the contributions from each protein residue $\Delta E_M \left(= E_M^{Ex} - E_M^G \right)$, where E_M^{Ex} and E_M^G are the interaction energies in the excited and ground states, respectively. We evaluated the interaction energy as the classical coulombic atomic charge interaction between $OxyLH_2$ and the surrounding molecules. The atomic charges of $OxyLH_2$ in the ground Q_A^G and the excited Q_A^{Ex} state were obtained by electrostatic potential fitting[30] using HF and CIS wave functions, respectively. Those of the residues Q_B were taken from the AMBER94 force-field. The result of the decomposition analysis is shown in Figure 6. The positive and negative contributions indicate spectral blue- and red-shifts in the fluorescence, respectively. It is clear that Arg218, His245, and AMP (phosphate group) give dominant contributions. Arg218 and AMP cause the blue-shifts of 0.29 and 0.21 eV, respectively, and His245 contributes to the red-shift of 0.17 eV. Overall, net ΔE becomes positive, and the surrounding residues and AMP lead to the

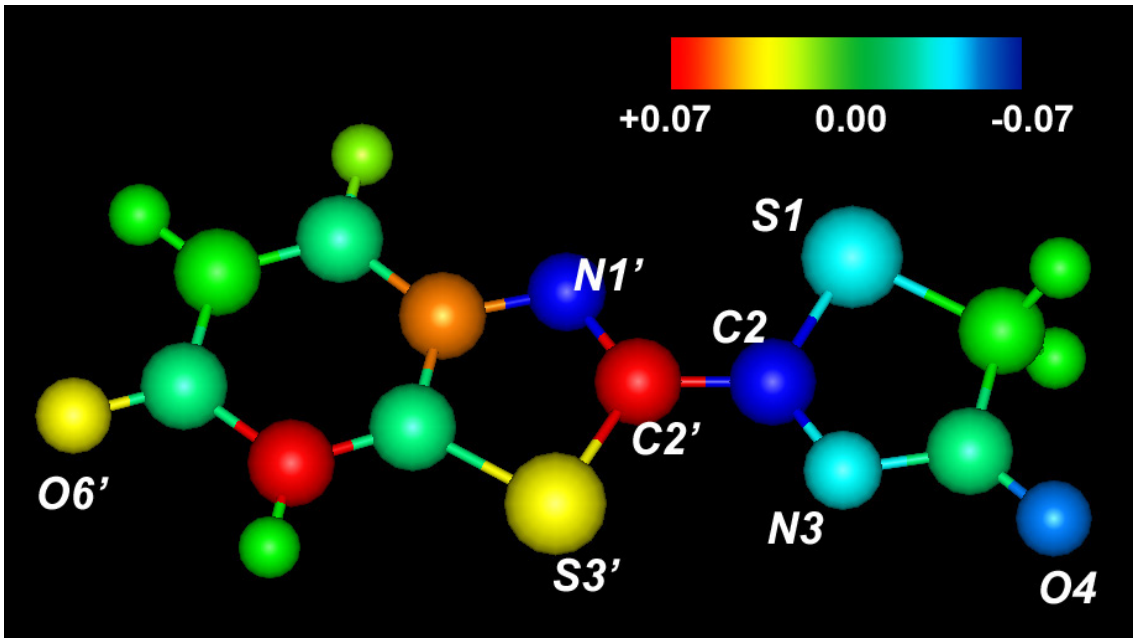


Figure 5. Difference in Mulliken atomic charge between the ground and the first excited state of $OxyLH_2$ (keto-s-trans form), $Q_A^{Ex} - Q_A^G$. The color of the atoms shows the change in the atomic charges.

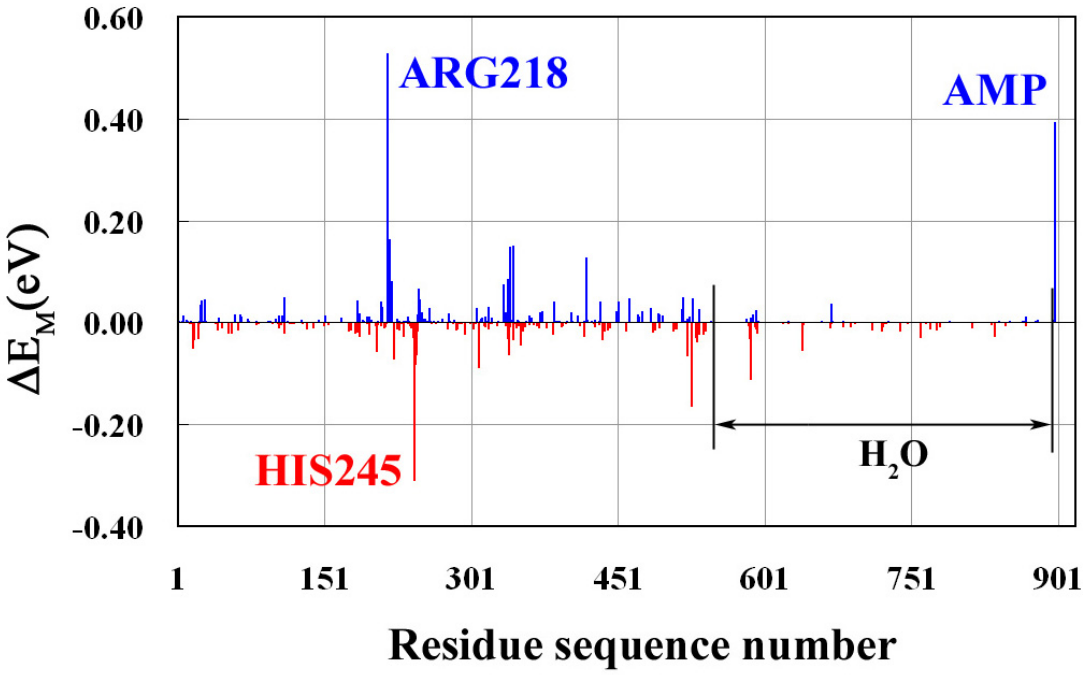


Figure 6. Decomposition analysis of ΔE (= $E^{Ex} - E^{G}$), difference in interaction energy between the ground and the excited state). ΔE was decomposed into contributions from the protein residues and from AMP, ΔE_M. Positive and negative contributions indicate spectral blue- and red-shifts in fluorescence, respectively.

Table 3. Emission (fluorescence) energies of $OxyLH_2$ in the keto-*s*-trans(-1) form in the gas phase and protein environment

Calc.	Environment	QM region	Geometry[a]	Emission energy / eV	
				SAC-CI	Exptl.
1	in Gas phase	$OxyLH_2$	Gas	1.97	
2		$OxyLH_2$	A-a	1.73	
			A-b	1.58	
3	in Protein	$OxyLH_2$	A-a	2.02	
			A-b	1.82	
4		+ ARG218	A-a	2.07	
			A-b	1.90	
5		+ HIS245	A-a	2.22	
			A-b	2.01	
6		+ Phosphate	A-a	2.33	2.23[b]
			A-b	2.08	
7		+PHE247, GLY341, LEU342, SER347, ALA348, HOH45	A-a	(2.24)[c]	2.23[b]
			A-b	(1.95)[c]	

[a] "Gas" denotes geometry optimized in the gas phase. For structures "A-a" and "A-b", see computational details. [b] Bioluminescence emission maxima for *Photinus pyralis* wild-type at pH 8.6.[9] [c] Energy correction due to the QM effect from the additional residues was estimated with eq. 2 (see text).

blue-shift in the emission energy. However, since these contributions seem to be too large to be treated by the electrostatic interactions alone, we treated Arg218, His245, and the phosphate group of AMP by the QM calculations with the SAC-CI method.

Table 3 shows the fluorescence energy calculated by the SAC-CI method using various computational models. The best calculation includes Arg218, His245, and the phosphate group, and is referred to as Calc. 6. Models A-a and A-b in Calc. 6 gave the excitation energies of 2.33 and 2.08 eV, respectively, as shown in Table 3. These values are close to the experimental emission energy (2.23 eV). The excitation is essentially one-electron

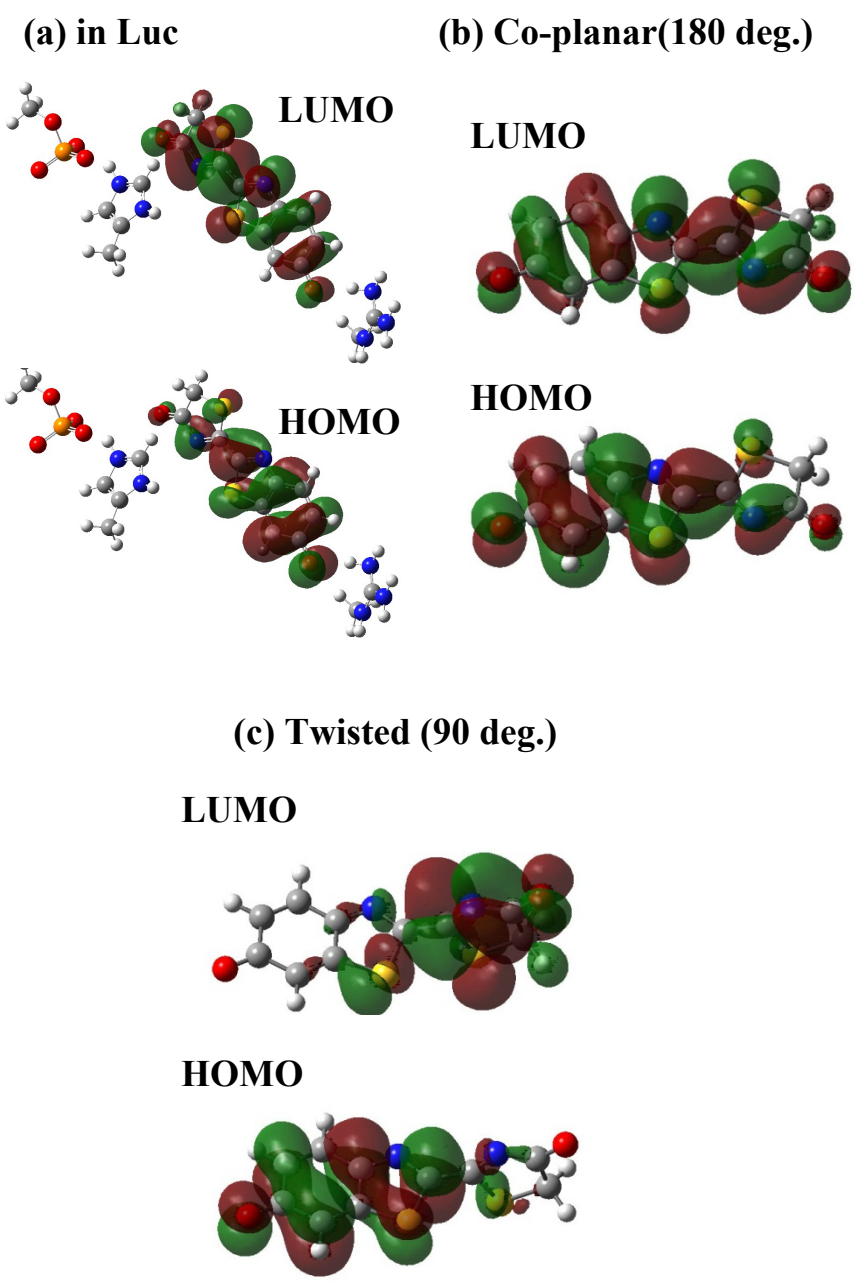


Figure 7. Highest-occupied canonical molecular orbital (HOMO) and lowest-unoccupied canonical molecular orbital (LUMO) of $OxyLH_2$ in (a, b) coplanar and (c) twisted structures. Red, blue, yellow, gray, and white balls denote O, N, S, C, and H atoms, respectively.

Table 4. Protein effect on emission energy analyzed by the SAC-CI calculations given in Table 3. See text for detailed explanations

Effect on the emission energy estimated by the series of SAC-CI calculations		Geometry[a]	Shift in emission energy (eV)
Total protein effect	Δ(Calc 6-Calc 1)	A-a	+0.36 (+0.27)[f]
		A-b	+0.11 (-0.02)[f]
Structure of $OxyLH_2$	Δ(Calc 2-Calc 1)	A-a	-0.24
		A-b	-0.39
Total environmental effect	Δ(Calc 6-Calc 2)	A-a	+0.60 (+0.51)[f]
		A-b	+0.50 (+0.37)[f]
Classical coulombic effect[b]	Δ(Calc 3-Calc 2)	A-a	+0.29
		A-b	+0.24
Quantum mechanical effect[c]	Δ(Calc 6-Calc 3)	A-a	+0.31 (+0.22)[f]
		A-b	+0.26 (+0.13)[f]
Only ARG218 in QM region[d]	Δ(Calc 4-Calc 3)	A-a	+0.05
		A-b	+0.08
ARG218 and HIS245 in QM region[e]	Δ(Calc 5-Calc 3)	A-a	+0.20
		A-b	+0.19

[a] "A-a" and "A-b" denote the structures A-a and A-b, respectively (see computational details). [b] All of the residues and AMP were treated by the point-charge model. [c] Arg218, His245, and the phosphate group of AMP were treated by the QM method. The rest were treated by the point-charge model. [d] Only Arg218 was included in the QM region in the SAC-CI calculation. The rest of the residues and AMP were treated by the point-charge model. [e] Arg218 and His245 were included in the QM region in the SAC-CI calculation. The rest of the residues and AMP were treated by the point-charge model. [f] The number in parentheses was obtained using Calc. 7 instead of Calc. 6 in Table 3. The QM effects from some other residues were evaluated at CIS level.

transition from HOMO to LUMO, and these orbitals are clearly localized within $OxyLH_2$, as shown in Figure 7a. This result suggests that the keto-form of $OxyLH_2$ can produce yellow-green luminescence. It agrees with the experiment by Branchini et al.,[9] which showed that yellow-green emission was produced even by a keto-constrained $OxyLH_2$ analog. It is not necessary to assume keto-enol tautomerism to explain the blue-shift of the emission wavelength. We also considered the possibility of the enol forms (see subsection of 'microenvironmental mechanism').

Next, the origin of the blue-shift was analyzed by comparing several SAC-CI calculations using different computational models. Table 4 summarizes the methods of the analyses and the results. The reference gas-phase SAC-CI calculation (Calc. 1) gives an excitation energy of 1.97 eV. This calculation used the optimized geometry in the gas phase, and no environmental effect was included. In Calc. 2, all of the surrounding molecules were removed from the computational models used in Calc. 6. Comparison between Calc. 2 and Calc. 1 gives the chromophore structural effect on fluorescence energy. The excitation energies obtained were 1.73 and 1.58 eV for models A-a and A-b, respectively. The structural constraint in the protein environment caused red-shifts in the fluorescence by 0.24 and 0.39 eV, respectively. This is because $OxyLH_2$ loses its co-planarity in the protein environment. Figure 8 shows the potential energy curves along the rotation of C2-C2' bond of $OxyLH_2$ (keto-s-trans(-1) form) in the gas phase. Since the energy of the ground state becomes more unstable than that of the first excited state, the C2-C2' rotation reduces the emission energy.

Comparison between Calc. 2 and Calc. 6 corresponds to the environmental effect on the emission energy caused by the coulombic interaction between $OxyLH_2$ and the surroundings. This environmental effect leads to a marked blue-shift in fluorescence energy of 0.60 and 0.50

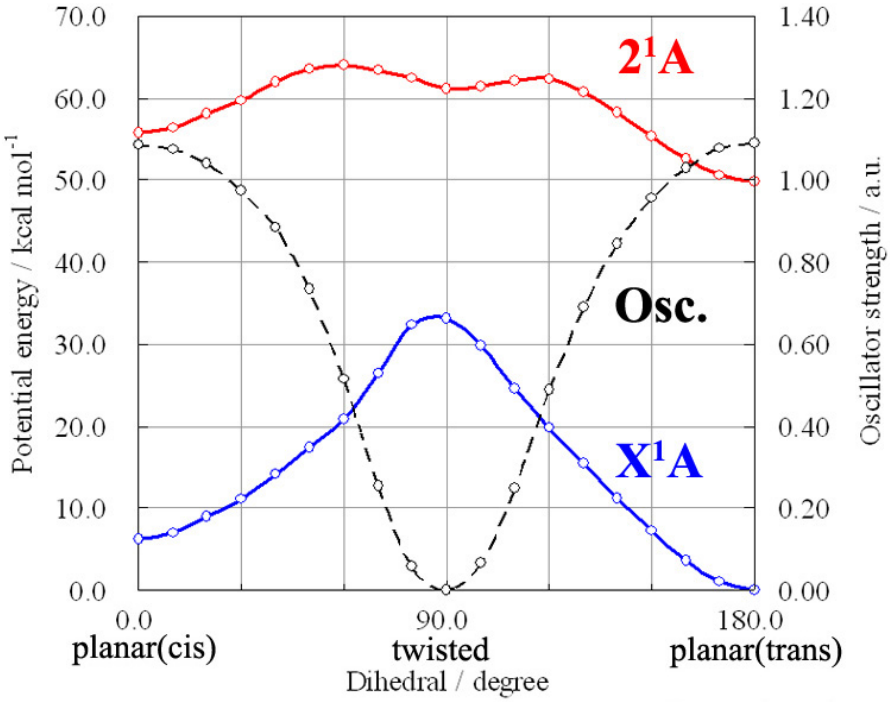


Figure 8. Potential energy curves along the rotation of the C2-C2' bond of $OxyLH_2$ (keto-s-trans(-1) form) in the gas phase. The other geometrical parameters were optimized at each point by CIS calculations for the first excited state. Oscillator strength in au (dashed line) is also shown.

eV in models A-a and A-b, respectively. To understand the origin of the shift, we performed further analyses using several SAC-CI results. In Calc. 3, all of the surroundings were treated by the point-charge model. The calculated emission energies were 2.02 and 1.82 eV for models A-a and A-b, respectively. Classical interaction contributes to the blue-shift, but the amounts of the shifts (0.29 and 0.24 eV for A-a and A-b, respectively) are smaller than those obtained with Calc 6. This indicates that the classical treatment can not sufficiently describe the interactions between $OxyLH_2$ and the surroundings.

Next, Arg218 was included in the QM region (Calc. 4). The calculated emission energies increased slightly to 2.07 and 1.90 eV for models A-a and A-b, respectively. We also included His245 in the QM region. The calculated emission energies became 2.22 and

2.01 eV, respectively. Inclusion of the phosphate group increased the excitation energies to 2.33 and 2.13 eV, respectively. The QM effect of Arg218, His245, and the phosphate group was estimated to be 0.05/0.08, 0.15/0.11, and 0.11/0.12 eV for model A-a/A-b, respectively. Thus, the QM effect from these residues also produces a spectral blue-shift of the fluorescence energy.

To investigate the QM effect of other neighboring residues and water molecules, we additionally included Phe247, Gly341, Leu342, Ser347, Ala348, HOH45 into the QM model (see Figure 4a,b for the computational model). Since SAC-CI calculation of such a large QM model is practically difficult at present, we applied ONIOM-like estimation[31] at CIS/D95(d) level.

$$EE(\text{large},\text{SAC-CI+CIS}) = \{EE(\text{large},\text{CIS}) - EE(\text{small},\text{CIS})\} + EE(\text{small},\text{SAC- CI}) \quad (2)$$

The correction obtained by this scheme was -0.09 and -0.13 eV for model A-a and model A-b, respectively. The estimated emission energies became 2.24 and 1.95 eV, respectively. Even though this correction is taken into account, the analysis in Table 4 shows that the total environmental effect gives a marked blue-shift in the emission energy: +0.51 and +0.37 eV for model A-a and model A-b, respectively.

2.3.3 On the possibility of the TICT state mechanism: potential energy surface of the first excited state of $OxyLH_2$

We now discuss the possibility of the TICT state as the emitting state. Figure 8 shows the gas-phase potential energy curves for the C2-C2' rotation (Figure 1c) calculated at the SAC-CI/D95(d)//CIS/D95(d) level. The calculated oscillator strength (emission intensity) is also shown.

The character of the first excited state is HOMO-LUMO transition. In the co-planar structure, both HOMO and LUMO are delocalized over the entire molecule, as shown in Figure 7b. In the twisted structure, however, the HOMO and LUMO are clearly localized on the benzothiazolyl ring and the thiazolinone ring, as shown in Figure 7c. These orbitals are very similar to the HOMO and LUMO of benzothiazole and thiazole, respectively. Since the first excited state in the twisted structure is a one-electron transition from HOMO to LUMO, the nature of the state is a charge-transfer state. The ground-state electronic structure is dominated by the closed-shell determinant. The coefficient of the double excitation operators is less than 0.1 in the SAC wave function.

On the potential surface of the first excited state, there are three stationary points: co-planar structures (0° and 180°) and a twisted structure (91°), both at the SAC-CI and CIS levels. The twisted structure is less stable than the co-planar structures. The energy barriers toward the co-planar structures are very small (less than 4 kcal/mol) in the first excited state. In contrast, the energy barriers to the twisted structure are 12.6 and 8.3 kcal/mol from the trans and the cis side, respectively. In addition, the oscillator strength for the fluorescence from the first excited state becomes close to zero in the twisted structure but is very large in the co-planar structures, as shown in Figure 8.

Based on our present results, even if $OxyLH_2$ became the TICT excited state during the photochemical processes, it could relax to a co-planar structure due to the small energy barrier. In addition, the fluorescence intensity is low in the twisted charge-transfer state. Therefore, the emitting state of $OxyLH_2$ should be a co-planar structure, and the TICT state mechanism is unrealistic.

2.3.4 On the micro-environmental mechanism

The hydrogen bonds between $OxyLH_2$ and the local environment have been implicated in the mechanism of the color-tuning in the firefly luminescence.[13,14] In this section, we discuss how this mechanism is relevant to yellow-green emission.

Firstly, the possibility of the enol form[13,14] was considered. Although Branchini and co-workers showed that keto-constrained $OxyLH_2$ emits green-yellow luminescence,[9,10] an enol transformation due to the local environment cannot be excluded in the case of the actual keto-$OxyLH_2$. We performed the SAC-CI calculations with CIS/6-31G* optimized geometry for enol-s-trans(-1) and enol-s-trans(-2) inside Luc. In both the enol-s-trans(-1) and the enol-s-trans(-2) structures, a crystal water molecule (HOH128) was located between the enol-$OxyLH_2$ and the phosphate group. In the enol-s-trans(-2) structure, the enol form was deprotonated, and the proton was transferred to the phosphate group. The fluorescence energy and energy profile are shown in Figure 9a, together with the optimized structures. The SAC-CI fluorescence energies (data in the parentheses) of keto-s-trans(-1), enol-s-trans(-1), and enol-s-trans(-2) in Luc were 2.33, 2.29, and 2.21 eV, respectively. All of them are close to the experimental value (2.23 eV). However, the total energy of the first excited state of the enol-s-trans(-1) and enol-s-trans(-2) structures are 19.8 and 34.2 kcal/mol higher than that of the keto-s-trans(-1) structure, respectively. Although the present CIS calculations are not quantitative, these energy differences are large enough for us to conclude that the enol transformation is energetically unfavorable in the Luc environment.

We now discuss the protonation state of the O6' atom[13,14] in the benzothiazoryl ring of $OxyLH_2$. We examined another protonation state in which Arg218 is neutral and the proton is transferred to $OxyLH_2$ (Figure 1d). The SAC-CI/D95(d) calculation was performed at the CIS/6-31G* optimized geometry. As shown in Figure 9b, the calculated fluorescence

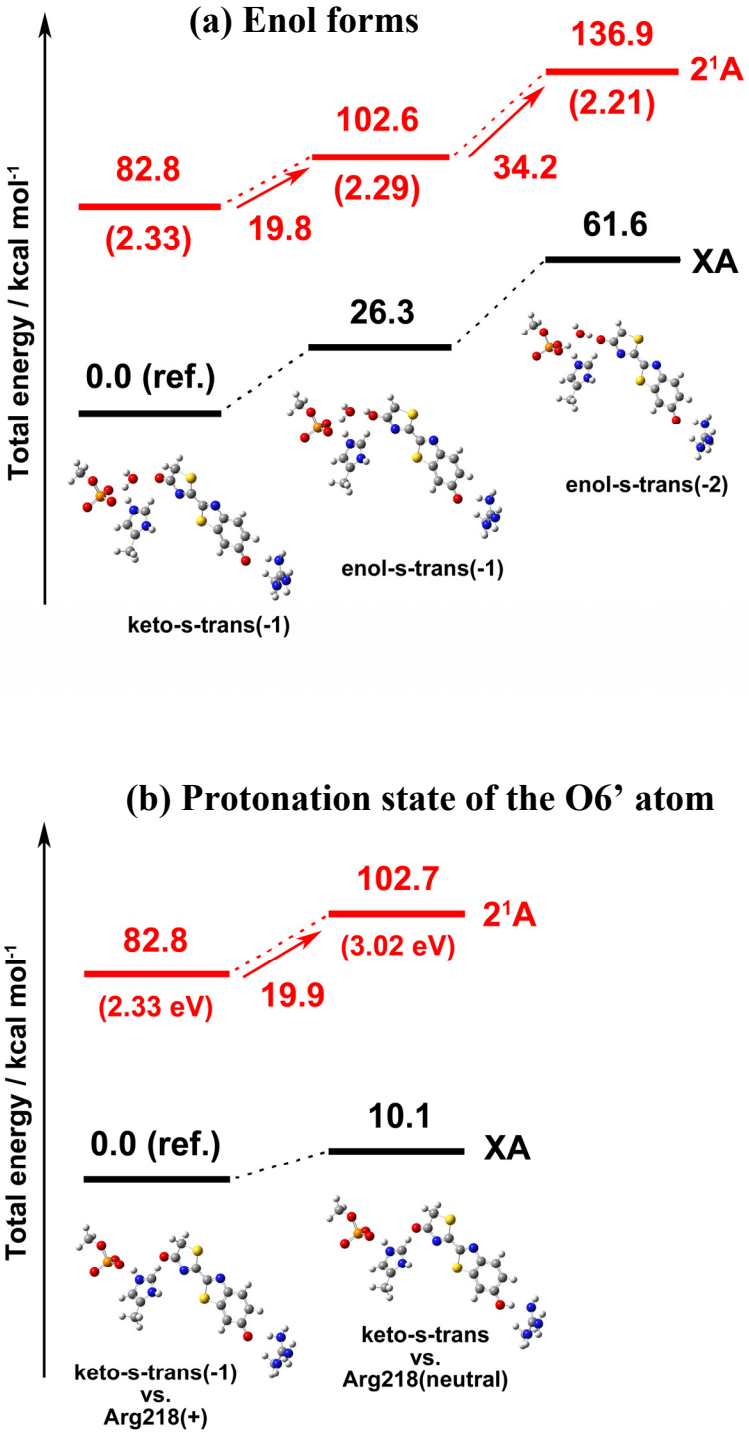


Figure 9. (a) Comparison of the potential energy and emission energy (in parenthesis) of the keto and enol forms in the Luc environment; (b) comparison of the potential energy and emission energy (in parenthesis) of the two protonation states.

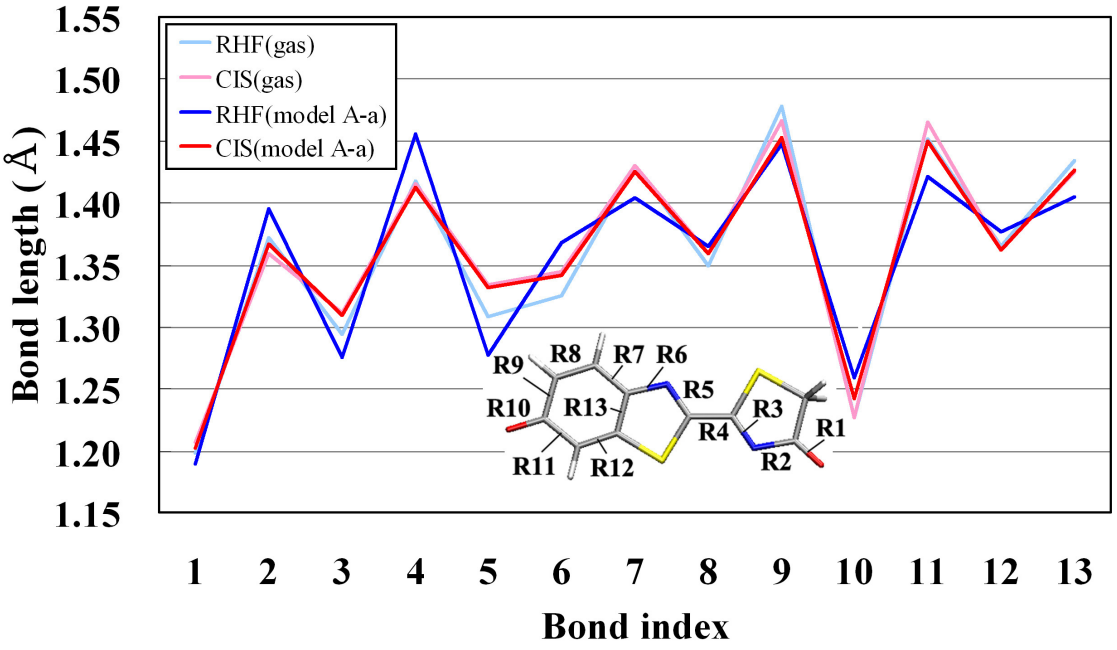


Figure 10. Bond lengths of the ground and the first excited states of keto-OxyLH$_2$ in the Luc environment (A-a) and in the gas phase optimized by HF and CIS methods.

energy was 3.02 eV, which is about 0.8 eV higher than the experimental value. In addition, the total energy evaluated at the CIS/6-31g* level was 20.2 kcal/mol higher than that of the keto-s-trans(-1) system. Therefore, we conclude that the neutral keto-OxyLH$_2$ (ketos-s-trans form) is not the emitting state in bioluminescence.

2.3.5 On the resonance-based mechanism

The mechanism originally reported was on the resonance structure in the excited state.[10] Here, we discuss how the resonance structures in both the ground and excited states affect the emission energy of keto-OxyLH$_2$. Since resonance structures are components of a wave function, the resonance structures characterize the properties of the state (including geometry). The optimized bond lengths of OxyLH$_2$ (keto-s-trans(-1)) are compared in Figure 10. The calculations were performed at RHF/6-31g* and CIS/6-31g* levels for the ground and the

excited state, respectively. The bond lengths for the first excited state in Luc are very close to those in the gas phase (the RMS deviation is 0.007 Å). This clearly indicates that the Luc environment leads to only small changes in the resonance structure in the excited state. In this sense, the original concept of the resonance structure-based mechanism differs from the actual mechanism of yellow-green luminescence.

In contrast, the bond length of the ground state is influenced by the Luc environment. The RMS deviation in bond length between the gas phase and Luc is 0.028 Å, and the maximum deviation amounts to 0.05 Å at R3, R4, and R5 (see Figure 10 for a definition of the bond labels). The bond-length alternation becomes marked in the Luc environment, indicating that the resonance structure in the ground state is more like the keto structure (left in Figure 1e). The ground-state relaxation energy from the excited-state geometry to the ground-state one is estimated to be 0.20 eV in Luc and 0.05 eV in the gas phase at the RHF/6-31g* level of calculation. However, such a structural change in the ground state enlarges the Stokes shift. To explain the blue shift in the emission energy, other blue-shift contributions (the protein electrostatic effect) should be larger than the red-shift effect of Stokes shift. Therefore, the structural change between the ground and excited states cannot explain the blue-shift in the Luc environment.

2.4 Conclusions

The yellow-green luminescence from firefly Luc of the North American firefly (*P. pyralis*) has long been ascribed to the enol form of $OxyLH_2$.[7] Recently, however, Branchini et al. showed that 5,5-dimethylluciferin constrained to the keto form can produce yellow-green emission in Luc.[10] In the present study, we theoretically determined the structure of the Luc-$OxyLH_2$ complex in the excited state: this is possible only with the

theoretical method, since it is the geometry in the excited state. The SAC-CI calculations were performed to calculate the fluorescence energies at the obtained structures.

The present SAC-CI results consistently explained the colors of firefly fluorescence produced by both chemiluminescence and bioluminescence. Depending on the environment, the emission energy shows a marked shift. The keto form of $OxyLH_2$ produced red-light chemiluminescence in solution (1.97 eV in experimental and 2.10 eV in SAC-CI theory). However, in the environment of Luc, the interaction with the nearby protein residues and AMP and the protein electrostatic effect causes the emission energy of the keto form shift markedly to a higher-energy region corresponding to green light (2.23 eV experimental and 2.08–2.33 eV in SAC-CI theory). The detailed analysis indicates that Arg218, His245, and the phosphate group of AMP play important roles in the color tuning mechanism, together with the protein electrostatic effect in the Luc environment.

Alternative mechanisms proposed in the previous studies were also examined but were not supported. (1) The TICT state[11] is meta-stable on the potential surface. However, it can easily relax to a co-planar structure due to the small activation barriers (around 2–4 kcal/mol). (2) The emission energy of enol-$OxyLH_2$ is very close to that of the keto form. However, the keto-enol transformation is difficult, since the enol forms are energetically unstable in Luc environment. (3) The protonation state of the O6' atom significantly affects the emission energy, as described previously.[14] However, proton transfer from Arg218 to keto-$OxyLH_2$ is energetically unfavorable. In addition, the emission energy is far above the observed value. (4) The resonance structures of $OxyLH_2$[10] in the ground and the excited state were also investigated. The resonance structure in the excited state would not be sensitive to the local environment.

Last, we note that to study the color-tuning mechanism in chemiluminescence and

bioluminescence, the theory used must be reliable enough for reproducing the excitation and luminescent energies. Otherwise, we can never do reliable analysis. In this sense, the SAC-CI method was very useful in the present study. For analyzing the principles and mechanisms behind the complex phenomena, the theory and the method used as a tool must be reliable enough, otherwise it may lead to erroneous conclusion.

2.5 Appendix

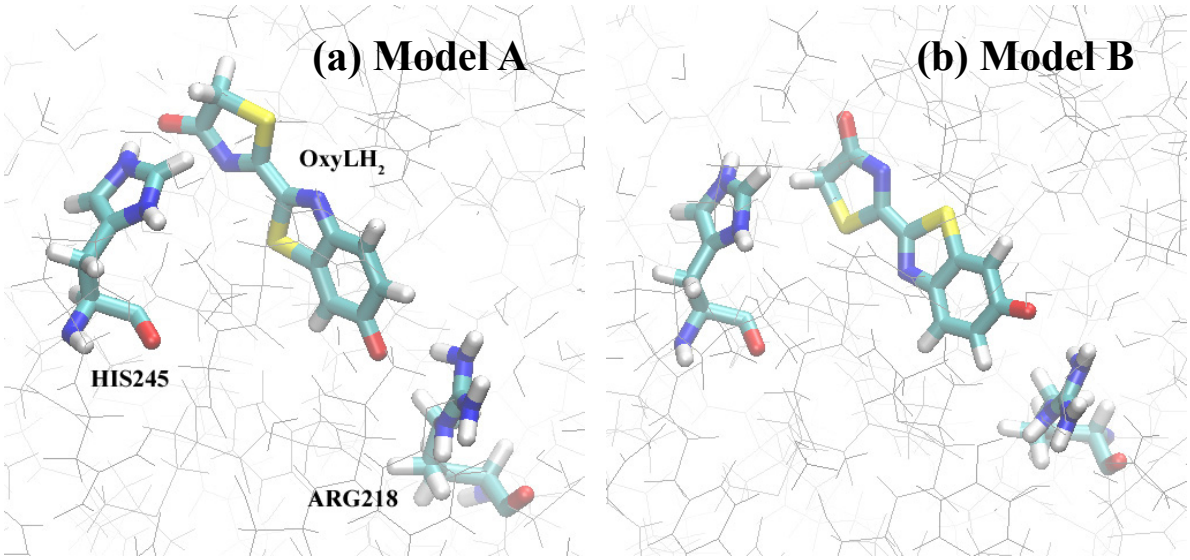


Figure A1. Snap shots taken from MD simulations for (a) model A and (b) model B.

Table A1. The bond distances optimized using CIS and SAC-CI methods.

Bond indices[a]	keto-*s*-trans(-1)			enol-*s*-trans(-1)		
	CIS/6-31g*	SAC-CI/6-31g*	Δ	CIS/6-31g*	SAC-CI/6-31g*	Δ
1	1.203	1.208	0.005	1.353	1.356	0.003
2	1.358	1.363	0.005	1.328	1.330	0.002
3	1.308	1.306	-0.002	1.333	1.333	0.000
4	1.410	1.430	0.020	1.406	1.417	0.011
5	1.331	1.311	-0.020	1.337	1.319	-0.018
6	1.343	1.377	0.034	1.342	1.367	0.025
7	1.427	1.415	-0.012	1.429	1.416	-0.013
8	1.353	1.370	0.017	1.359	1.371	0.012
9	1.463	1.458	-0.005	1.467	1.459	-0.008
10	1.221	1.238	0.017	1.230	1.237	0.007
11	1.460	1.457	-0.003	1.461	1.460	-0.001
12	1.359	1.380	0.021	1.363	1.376	0.013
13	1.422	1.408	-0.014	1.428	1.416	-0.012
14	1.766	1.762	-0.004	1.766	1.762	-0.004
15	1.763	1.768	0.005	1.770	1.773	0.003
16	1.776	1.782	0.006	1.756	1.762	0.006
17	1.806	1.812	0.006	1.742	1.750	0.008
18	1.541	1.544	0.003	1.372	1.371	-0.001
RMS			0.014			0.010

[a] See Figure A1.

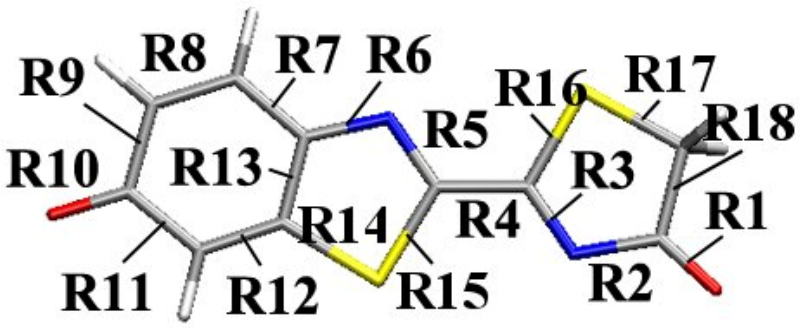


Figure A2. Bond indices corresponding the notation in Table A1.

Table A2. CIS emission energies of OxyLH2 calculated with the "tight ion-pair" and "solvent-separated ion-pair" models

Molecule	Solvation Model	Emission Energy / eV
keto-*s*-trans(-1)	K^+ / PCM (DMSO)	3.10
	DMSO + K^+ / PCM (DMSO)	3.07
enol-*s*-trans(-1)	K^+ / PCM (DMSO)	3.04
	DMSO + K^+ / PCM (DMSO)	3.00
enol-*s*-trans(-1)'	K^+ / PCM (DMSO)	3.11
	DMSO + K^+ / PCM (DMSO)	3.06
enol-*s*-trans(-2)	K^+ / PCM (DMSO)	3.01
	DMSO + K^+ / PCM (DMSO)	2.97

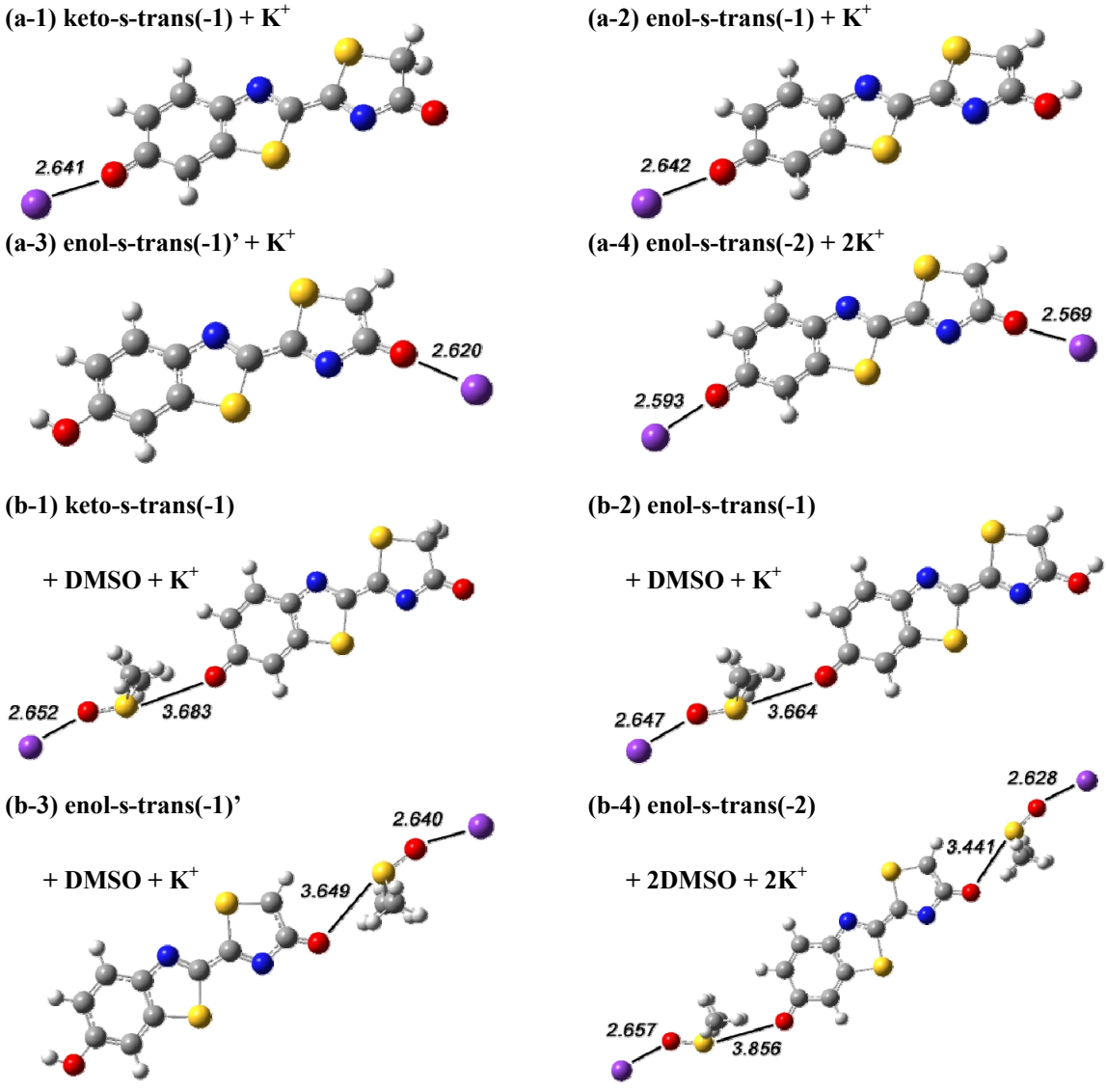


Figure A3. Optimized structures of the excited state of $OxyLH_2$ in K+/DMSO solution.

References

(1) (a)Wood, K. V., *Photochem. Photobiol.* **1995**, *62*, 662. (b)Wood, K. V.; Lam, Y. A.; Seliger, H. H.; McElroy, W. D., *Science* **1989**, *244*, 700. (c)Viviani, V. R; Bechara, E. J. H., *Photochem. Photobiol.* **1995**, *62*, 490. (d)Viviani, V. R., *Cell. Mol. Life Sci.* **2002**, *59*, 1833. (e)Ugarova, N. N., *Biolum. Chemilum.* **1989**, *4*, 406.

(2) Kricka, L. J., *Methods Enzymol.* **2000**, *305*, 333.

(3) Deluca, M., *Adv. Enzymol.* **1976**, *44*, 37.

(4) (a)Deluca, M.; Leonard, N. J.; Gates, B. J.; McElroy, W. D., *Proc. Natl. Acad. Sci. U. S. A.* **1973**, *70*, 1664. (b)DeLuca, M.; McElroy, W. D., *Biochemistry* **1974**, *13*, 921.

(5) McCapra, F., *J. Chem. Soc. Chem. Commun.* **1977**, 946; Koo, J.-Y.; Schmidt, S. P.; Schuster, G. B., *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 30; Schuster, G. B., *Acc. Chem. Res.* **1979**, *12*, 366.

(6) White, E. H.; Rapaport, E.; Seliger, H. H.; Hopkins, T. A., *Bioorg. Chem.* **1971**, 92.

(7) White, E. H.; Rapaport, E.; Hopkins, T. A.; Seliger, H. H., *J. Am. Chem. Soc.* **1969**, *91*, 2178.

(8) White, E. H.; Steinmetz, M. G.; Miano, J. D.; Wildes, P. D.; Morland, R., *J. Am. Chem. Soc.* **1980**, *102*, 3199.

(9) Branchini, B. R.; Murtiashaw, M. H.; Magrar, R. A.; Portier, N. C.; Ruggiero, M. C. Stroh, J. G., *J. Am. Chem. Soc.* **2002**, *124*, 2112.

(10) Branchini, B. R.; Southworth, T. L.; Murtiashaw, M. H.; Magyer, R. A.; Gonzalez, S. A.; Ruggiero, M. C.; Stroh, J. G., *Biochemistry* **2004**, *43*, 7255.

(11) McCapra, F.; Gilfoyle, D. J.; Young, D. W.; Church, N. J.; Spencer, P., *Bioluminescence and Chemiluminescence. Fundamentals and Applied Aspects*; Campbell, A. K.; Kricka, L. J.; Stanley, P. E., Eds.; John Wiley & Sons: Chichester, 1994, pp. 387.

(12) DeLuca, M., *Biochemistry* **1969**, *8*, 160.

(13) (a)Ugarova, N. N.; Brovko, L. Y., *Luminescence* **2002**, *17*, 321. (b)Gandelman, O. A.; Brovko, L. Y.; Ugarova, N. N.; Chikishev, A. Y.; Shkurimov, A. P., *J. Photochem. Photobiol. B* **1993**, *19*, 187.

(14) Orlova, G.; Goddard, J. D.; Brovko, L. Y., *J. Am. Chem. Soc.* **2003**, *125*, 6962.

(15) Conti, E.; Franks, N. P.; Brick, P., *Structure* **1996**, *4*, 287.

(16) (a)Branchini, B. R.; Magyar, R. A.;M urtiashaw, M. H.; Anderson, S. M.; Zimmer, M., *Biochemistry* **1998**, *37*, 15311. (b)Branchini, B. R.; Magyar, R. A.; Murtiashaw, M. H.; Anderson, S. M.; Helgerson, L. C.; Zimmer, M., *Biochemistry* **1999**, *38*, 13223. (c)Branchini, B. R.; Southworth, T. L.; Murtiashaw, M. H.; Boije, H.; Fleet, S. E., *Biochemistry* **2003**, *42*, 10429.

(17) Branchini, B. R.; Magyar, R. A.; Murtiashaw, M. H.; Portier, N. C., *Biochemistry* **2001**, *40*, 2410..

(18) Nakatsuji, H.; Hirao, K., *J. Chem. Phys.* **1978**, *68*, 2053.

(19) (a)Nakatsuji, H., *Chem. Phys. Lett.* **1978**, *59*, 362. (b)Nakatsuji, H., *Chem. Phys. Lett.* **1989**, *67*, 329. (c) Nakatsuji, H., *Chem. Phys. Lett.* **1989**, *67*, 334.

(20) Nakatsuji, H., *Computational Chemistry, Revies of Current Trends*; Leszczynski, J., Ed.; World Scientific: Singapore, 1996, Vol. 2, pp. 62.

(21) (a) Casida, M. E.; Jamorski, C.; Casida, K. E.; Salahub, D. R. *J. Chem. Phys.* **1998**, *108*, 4439. Casida, M. E.; Salahub, D. R. *J. Chem. Phys.* **2000**, *113*, 8918. (b) Appel, F.; Gross, E. K. U.; Burke, K. *Phys. Rev. Lett.* **2003**, *90*, 043005. (c) Hsu, C.-P.; Hirata, S.; Head-Gordon, M. *J. Phys. Chem. A* **2001**, *105*, 451-458. (d) Dreuw, A.; Weisman, J. L.; Head-Gordon, M. *J. Chem. Phys.* **2003**, *119*, 2943-2946. (e) Milet, A.; Korona, T.; Moszynski, R.; Kochanski, E. *J. Chem. Phys.* **1999**, *111*, 7727. (f) Fujimoto, K.; Hayashi,

S.; Hasegawa, J.; Nakatsuji, H. *J. Chem. Theory Comput.* **2007**, *3*, 605-618.

(22) Nakatsuji, H., *Chem. Phys.* **1983**, *75*, 425.

(23) (a)Miertus, S.; Scrocco, E.; Tomasi, J., *J. Chem. Phys.* **1981**, *55*, 117. (b)Miertus, S.; Tomasi, J., *J. Chem. Phys.* **1982**, *65*, 239. (c)Cossi, M.; Barone, V.; Cammi, R.; Tomasi, J., *J. Chem. Phys. Lett.* **1996**, *255*, 327. (d)Barone, V.; Cossi, M.; Mennucci, B.; Tomasi, J., *J. Chem. Phys.* **1997**, *107*, 3210.

(24) Dunning, T. H.; Hay, P. J., *Modern Theoretical Chemistry*; Schaefer, H. F., Ed.; Plenum: New York, 1976, Vol. 3, pp. 1.

(25) Nakatsuji, H.; Ushio, J.; Yonezawa, T., *Can. J. Chem.* **1985**, *63*, 1857.

(26) Pipek, J.; Mezey, P. G., *J. Chem. Phys.* **1989**, *90*.

(27) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., *J. Am. Chem. Soc.* **1995**, *117*, 5179.

(28) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, M.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin,

R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A., Gaussian03, Gaussian, Inc., Pittsburgh, 2003.

(29) Ponder, J. W.; Ren, P., *J. Phys. Chem. B* **2003**, *107*, 5933.

(30) Besler, B. H.; Merz Jr, K. M.; Kollman, P. A., *J. Comp. Chem.* **1990**, *11*, 431.

(31) F. Maseras and K. Morokuma, *J. Comp. Chem.* **1995**, *16*, 1170.

Chapter 3

Artificial Color Tuning of Firefly Luminescence: Theoretical Mutation by Tuning Electrostatic Interactions between Protein and Luciferin

3.1 Introduction

Firefly luciferase (Luc) is an oxygenaze enzyme which catalyzes the light-emitting reaction. Luc oxygenates firefly luciferin and produces oxyluciferin in the electronically excited state (see Figure 1a).[1-5] During the decay process to the ground state, keto-oxyluciferin ($oxyLH_2$) shows yellow-green emission.[1,2] The quantum yield of the luminescence is known to be very high.[6] Firefly Luc has successfully been applied to various fields of bio-imaging[2] to assay ATP generation,[7] gene expression,[8] and biosensors for environmental pollutants.[8] It is desirable to develop Luc showing different emission colors.[2] In particular, red-light emitting Luc variants are useful for the applications to mammalian cells which absorb shorter wavelength.[2,9]

Recently, a red emitting Ser284Thr mutant (615 nm, 2.02 eV) and a green emitting Val241Ile/Gly246Ala/Phe250Ser triple mutant (549 nm, 2.26 eV) of *Photinus pyralis* (*Ppy*) have been developed and showed favorable spectroscopic and kinetic properties.[10] A synergistic effect was observed in a mutagenesis study,[11] indicating that the active-site local structure and the steric effect were related to the spectral tuning.[11] Regarding *Luciola cruciata*, two red emitting mutants, Ile288Val (613 nm, 2.02 eV) and Ile288Ala (613 nm, 2.02 eV) were also produced.[12] According to the X-ray structural observations, the spectral shift correlates with the size of the side chain in the 288-th amino acid.[12] These successful

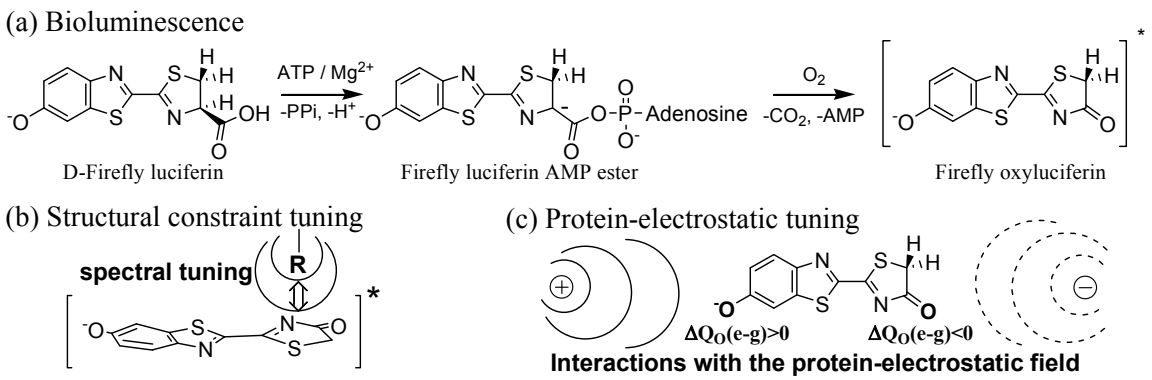


Figure 1. (a) Reaction scheme of the firefly-luciferin enzyme reaction. (b) Protein-electrostatic effect on the excited state of $OxyLH_2$. $\Delta Q_O(e\text{-}g)$ is the atomic-charge difference of oxygen atoms between the excited and ground states of $OxyLH_2$ ($\Delta Q_O(e\text{-}g) = Q_O(e) - Q_O(g)$).

examples are thus based on the steric constraint which controls the structural relaxation of the excited $oxyLH_2$ before the emission (Figure 1b).

In our previous study,[13] we analyzed the emission-color tuning mechanism of *Ppy* Luc, why red light in chemiluminescence and yellow-green in Luc. The $oxyLH_2$ structural distortion induced in the Luc environment reduces the emission energy by about 0.24-0.39 eV. The C-C dihedral angles between the two rings were ascribed as the main structural difference. This result indicated that the emission energies might be controlled by modifying the structural relaxations in the excited state of $oxyLH_2$, which supports the steric constraint strategy.[10-12]

We have also shown that the quantum and classical Coulombic interactions between $oxyLH_2$ and Luc increase the emission energy by about 0.50-0.60 eV. We performed quantum-theoretical calculations and clarified the crucial role of the protein-electrostatic effects on the emission energy. Since the charge distribution of $oxyLH_2$ changes upon the electronic transition, the emission energy is sensitive to the protein-electrostatic environment

as schematically shown in Figure 1c. This effect was shown to be a major factor for making the emission color to be yellow-green.[13] Therefore, it is a reasonable strategy to control the electrostatic interactions between $oxyLH_2$ and Luc proteins. Once the interactions between $oxyLH_2$ and amino acids are analyzed, we are able to identify several key amino acids which control the emission energy. It is therefore possible to propose new mutation experiments for the emission-spectral tuning.

For calculating emission energies, we used the Symmetry-Adapted Cluster-Configuration Interaction (SAC-CI) method.[14-17] The method has been applied to more than 150 molecules and established as an accurate electronic-structure theory for the ground and excited states.[18-20] In our recent study, we successfully reproduced the emission energies of $oxyLH_2$ observed in chemiluminescence and bioluminescence.[13] The method also successfully applied to analyze the color-tuning mechanisms in retinal proteins[21-24] and green-fluorescent proteins.[25,26]

The aim of the present study is to propose mutations to control the electrostatic interactions between $oxyLH_2$ and Luc for tuning the color of the firefly luminescence. First, the results of the decomposition analysis of the electrostatic interaction between $oxyLH_2$ and the surrounding molecules are described. Next, on the basis of the electrostatic strategy, our new proposals of the mutations are proposed. Finally, to numerically examine our proposals, we performed the SAC-CI emission-energy calculations, and the results of the theoretical mutation simulations are described.

3.2 Computational Details

The X-ray structure (1LCI[27]) and the working model proposed in the experimental studies[28-32] were used for constructing the $oxyLH_2$-Luc binding complexes. Molecular

dynamics calculations with AMBER94 force-field[33] were performed to find the stable structures. The force-field parameters for oxyLH$_2$ were designed for the first excited state.[13] When all of atoms in an amino acid were located more than 7.0 Å away from oxyLH$_2$, the atomic coordinates were frozen in the molecular dynamics (MD) calculations. We took three stable structures from the trajectory. These three had different AMP conformations. There were only minor differences in the other part of the structures. Starting from these three, we raised the temperature stepwise up to 300 K, in which total 60 ps MD simulations were performed. Then we performed 30 ps MD simulations at 300 K. After the MD calculations, the molecular mechanics (MM) energy minimizations were performed. The MM minimized structures were used for the initial guesses for the Configuration Interaction Singles (CIS) geometry optimizations.

In the CIS calculations, oxyLH$_2$, AMP, and amino acids in the first solvation layer (ARG218, HIS245, PHE247, GLY341, LEU342, SER347, ALA348, and the crystal water HOH45) were treated with the quantum mechanical (QM) method. The rest of the protein effect was included by the classical point-charge model defined by the MM parameters. Geometry of OxyLH$_2$ was optimized in the CIS calculations. The other atoms in the QM region were fixed to the MM optimized structures. The basis sets of 6-31g* were used for oxyLH$_2$, and the 6-31g set for the rest of the QM atoms.

Single-point SAC-CI calculations were performed at the optimized geometries to evaluate the fluorescence energies. OxyLH$_2$, His245, Arg218, and the phosphate group of AMP were treated quantum mechanically, and the rest was done by a classical point-charge model. All valence orbitals were included in the active space of the SAC-CI calculations. Perturbation selection was performed to select the double excitation operators.[34] The energy thresholds of 5.0×10^{-6} and 5.0×10^{-7} hartree (LevelTwo) were used for the ground and excited

states, respectively. For the basis sets, we used Huzinaga–Dunning double-zeta basis sets[35] augmented with the single-polarization d-function.[35] The calculated emission energy for the wild type, 2.23 eV, was in good agreement with the experimental one, 2.33 eV.[13]

In constructing the computational model, we trusted the X-ray structure. Comparing the structures available in the Protein Data Bank,[27,36] the luciferase structures was not so sensitive to the inhibitor binding in the active site.[36] In our previous study,[13] including the oxyluciferin in the binding site gave only small structural changes. One reason is that luciferase has open space in the binding site. If there is no inhibitor inside, the space is almost filled by water molecules.

The QM and MM calculations were performed using the *Gaussian* 03[37] and TINKER programs,[38-40] respectively.

3.3 Result and Discussions

3.3.1 Role of the protein-oxyLH_2 interactions behind the yellow-green emission of the wild type.

Since bioluminescences originate from the electronic transition, the charge distribution of the chromophore changes after the luminescence. In our previous study, the emission color tuning mechanism was ascribed to the increase/decrease in the charges of the carbonyl O atoms.[13] However, the transition is better characterized as an intramolecular charge transfer from the benzothiazole ring to the thiazole moiety. Figure 2 shows the orbital distributions of the lowest unoccupied molecular orbital (LUMO) and the highest occupied molecular orbital (HOMO). As seen in Figure 2b, the benzothiazole ring in the HOMO is more populated than the thiazole ring. In contrast, the LUMO shown in Figure 2a distributes slightly more in the thiazole moiety. Dipole moments clearly reflect the charge-transfer character as shown in

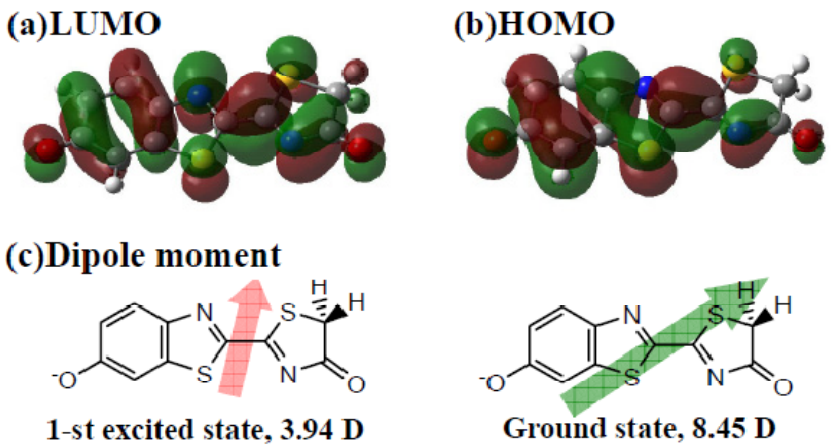


Figure 2. Orbital distributions of (a) LUMO and (b) HOMO. (c) Dipole moment of $OxyLH_2$ in the first excited and ground states.

Figure 2c. $OxyLH_2$ is polar in the ground state (8.45 D), but, in the excited state, the dipole moment is much smaller (3.94 D).

Due to the charge-transfer character, the emission energy of $oxyLH_2$ is influenced by the electrostatic field generated by the Luc environment, which were the major origin of the yellow-green emission in the wild type (WT). We analyzed the protein-electrostatic effect at the amino acids resolution as follows.

$$\begin{aligned}\Delta E &= \sum_{M}^{\text{Amino acids}} \Delta E_M = \sum_{M}^{\text{Amino acids}} \left\{ E_M^{Ex} - E_M^{G} \right\} \\ &= \sum_{M}^{\text{Amino acids}} \left\{ \sum_{B \in M} \sum_{A \in \text{OxyLH}_2} \frac{\left(Q_A^{Ex} - Q_A^{G} \right) Q_B}{r_{AB}} \right\}\end{aligned} \tag{1}$$

The ΔE is the difference in the interaction energy between the ground and excited states, and can be easily decomposed into the contribution from each amino acid, $\Delta E_M \left(= E_M^{Ex} - E_M^{G} \right)$. The E_M^{Ex} and E_M^{G} are the electrostatic interaction energies in the excited and ground states, respectively. We evaluated the interaction energy as the classical Coulombic interaction between $oxyLH_2$ and the surrounding molecules. The atomic charges of $oxyLH_2$ in the ground and the excited states, Q_A^{G} and Q_A^{Ex}, were obtained by the

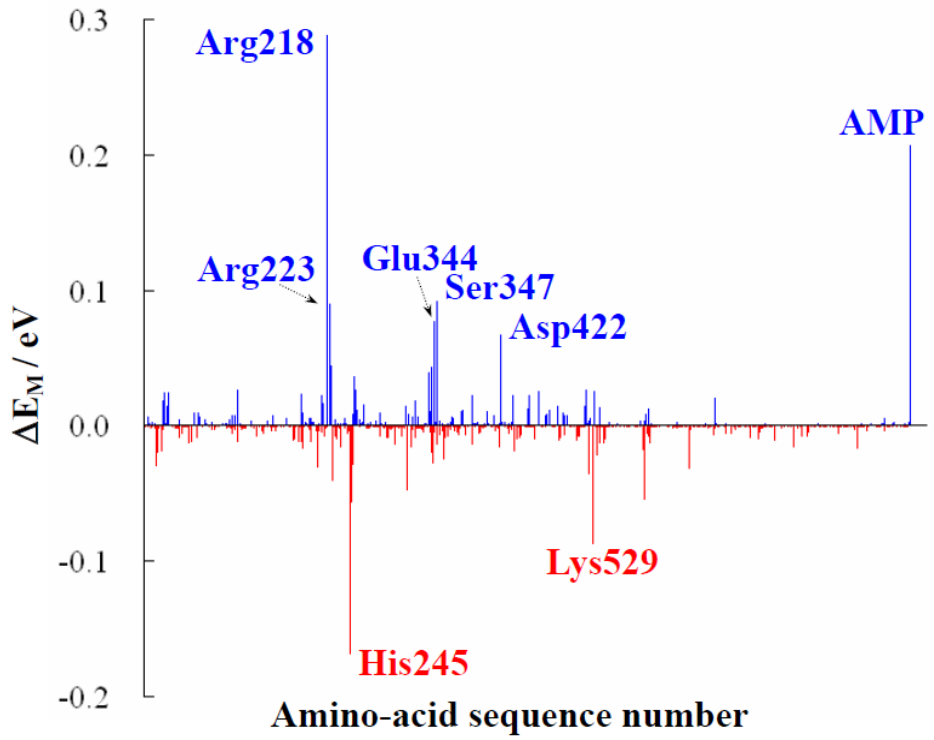


Figure 3. Decomposition analysis of ΔE ($=E^{Ex}-E^{G}$, difference of the interaction energies in the ground and the excited states). ΔE was decomposed into contributions from the protein residues and AMP (ΔE_M). Positive and negative contributions indicate spectral blue- and red-shifts in fluorescence, respectively.

electrostatic-potential fitting using HF and CIS wave functions, respectively. The Q_B for the atoms in Luc were taken from the AMBER94 force-field.[33]

The result of the decomposition analysis is shown in Figure 3. The blue- and red-shift contributions in the emission energy caused by each amino acid were indicated by the positive and negative regions of the diagram, respectively. It is clearly shown that Arg218, His245, and AMP (phosphate group) give the primary contributions. Arg218 and AMP cause the blue-shifts of 0.29 and 0.21 eV, respectively, and His245 contributes to the red-shift of 0.17 eV.[13] In a previous experimental mutation study, fluorescence energies of Arg218Ala and Arg218Gln mutants were observed to be 2.03 (611) and 2.04 eV (608 nm), respectively,

which were around 0.2 eV smaller than that of the wild type.[32] These experimental results qualitatively agree to our analysis, because the positively-charged residue at the 218-th position causes the blue shift. A Ser347Ala mutation red-shift around 0.17 eV)[30] also supports the present theoretical result.

In this study, we focus on the secondary important contributions from the amino acids relatively distant form the binding site: blue-shift contributions from Arg223, Glu344, and Asp422, and red-shift contribution from Lys529 as shown in Figure 3. The positions and structures of these amino acids were shown in Figure 4. Since Arg223 is positively charged and located close to Arg218, this amino acid contributes the positive electrostatic potential as shown in Figure 1c. In contrast, negatively charged Glu344 and Asp422 provide the negative electrostatic potential in the same way as the AMP phosphate group.

3.3.2 Artificial red- and yellow-emitting firefly luciferases: theoretical mutation simulations.

On the basis of the analysis described above, we point out some amino acids to obtain red-emitting firefly luciferases. As shown in Figure 3, Arg223, Glu344, and Asp422 have blue-shift contributions in the emission energies of the wild type. When they are replaced by a neutral residue as alanine, the blue-shift effect should disappear, and the emission energy is expected to show red shifts. Since the distances between oxyLH$_2$ and the target residues are around 7-12 Å, we assumed the structural rearrangements around the oxyLH$_2$ binding site causes minor effect. On the other hand, Arg218, His245, and Ser347 were excluded. Since these residues are in direct contact with oxyLH$_2$, the mutations would change the oxyLH$_2$ binding conformations. Once the primary binding-site residue was lost, it would be very difficult to find a reasonable binding structure between oxyLH$_2$ and Luc in computational

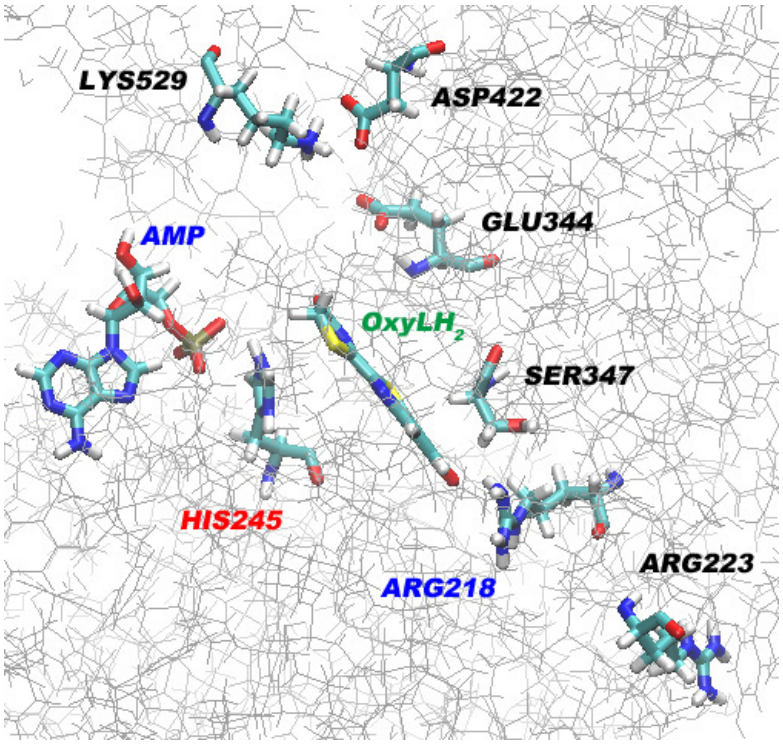


Figure 4. Amino acids relevant to the emission color tuning. The structure of the Luc-OxyLH_2 complex was optimized for the first excited state of OxyLH_2.

Table 1. Emission energies of the mutants calculated by the SAC-CI method.

System	Theoretical			Exptl.[b]
	SAC-CI	Spectral shift	Shifted[a,b]	
Wild type	2.33[c]	-	2.23(557)	2.23(557)[d]
R223A	2.27	-0.06	2.17(571)	
E344A	2.22	-0.11	2.12(585)	
D422A	2.25	-0.08	2.15(577)	
R223A/E344A/D422A	2.15	-0.18	2.05(605)	

[a] Emission energies were shifted by −0.1 eV, so that the emission energy of the wild type became equal to the experimental value. [b] Unit is in eV. The numbers of in parentheses are in nm unit. [c] Ref. 13. [d] Ref. 10.

point of view. The experimental mutations at Arg218, His245, and Ser347 were already reported.[29,30,32] Significant spectral shifts were observed in these mutants,[32] which qualitatively agree to our result of the electrostatic analysis. However, since the mutation, especially Arg218, also dropped the oxyLH$_2$ binding capability of the luciferase, the oxygenase catalytic reactions became less efficient than that in the wild type.[29,30,32]

Next we performed theoretical mutation simulations. The Arg223, Glu344, and Asp422 were replaced by alanines, and the structures of the amino acids were optimized by the MM calculations. The SAC-CI calculations were performed to evaluate the fluorescence energies of the mutants.

In Table 1, the calculated emission energies were summarized. In the Arg223Ala, Glu344Ala, and Asp422Ala mutants, the emission energies were calculated to be 2.27 eV, 2.22 eV, and 2.25 eV, respectively. The estimated spectral red-shifts are 0.06 eV, 0.11 eV, and 0.08 eV, respectively. The amounts of the shifts qualitatively agree with those obtained by the analysis as shown in Figure 3. Table 1 also showed the "Shifted" emission energies which were shifted by -0.10 eV so that the theoretical result of the wild type became equal to the experimental one. The expected emission energies of the Arg223Ala, Glu344Ala, and Asp422Ala mutants were 2.17 eV (571 nm), 2.12 eV (585 nm), and 2.15 eV (577 nm), respectively. Although the amounts of the shifts are not so significant, we expect yellow emissions from these single mutants.

We further examined a triple mutant, Arg223Ala/Glu344Ala/Asp422Ala. The amount of the estimated red-shift became 0.18 eV and the expected emission energy became 2.05 eV (605 nm), which is the reddish-orange emission.

To the best of our knowledge, none of the amino acids pointed out in the present study has been tested in previous experimental mutations. Since the targeted amino acids locate

relatively far from oxyLH$_2$ binding site, the mutations would not directly change the binding site. Therefore, we expect that the mutation effects proposed based on the present and previous studies[10-12] would provide additive effects in controlling the emission energy. We strongly hope that the present theoretical predictions are tested by real mutation experiments.

3.4 Conclusions

There are several successful firefly Luc mutants showing the red and green emissions.[10,12] The mechanism proposed in these studies was based on the steric constraint.[11,12] On the other hand, we found that the protein-electrostatic effect gave the largest contribution to the blue-shift in the emission energy.[13] In this study, we proposed three amino acids for the mutation experiments, Arg223, Glu344, and Asp422 in *Ppy* Luc.

Theoretical mutation simulations using the SAC-CI method showed that the Arg223Ala, Glu344Ala, and Asp422Ala mutations showed spectral-red shift of 0.06, 0.11, and 0.08 eV, respectively. A triple mutant, Arg223Ala/Glu344Ala/Asp422Ala, showed a larger red-shift of 0.18 eV. The emission energy expected is 2.05 eV (605 nm, reddish-orange). We thus wish that our proposal would add color variations in addition to the existing successful mutants.[10,12]

References

(1) White, E. H.; Rapaport, E.; Seliger, H. H.; Hopkins, T. A. *Bioorg. Chem.* **1971**, 92-122.

(2) Viviani, V. R. *Cell. Mol. Life Sci.* **2002**, *59*, 1833-1850.

(3) Isobe, H.; Takano, Y.; Okumura, M.; Kuramitsu, S.; Yamaguchi, K. *J. Am. Chem. Soc.* **2005**, *127*, 8667-8679.

(4) Chung, L. W.; Hayashi, S.; Lundberg, M.; Nakatsu, T.; Kato, H.; Morokuma, K. *J. Am. Chem. Soc.* **2008**, *130*, 12880-12881.

(5) Vico, L. D.; Liu, Y.-J.; Krogh, J. W.; Lindh, R. *J. Phys. Chem. A* **2007**, *111*, 8013-8019.

(6) Seliger, H. H.; McElroy, W. D. *Arch. Biochem. Biophys.* **1960**, *88*, 136-141.

(7) Campbell, A. K. *Chemiluminescences: Principles and Applications in Biology and Medicine.*, VHC, Chichester, 1988.

(8) Naylor, L. H. *Biochem. Pharm.* **1999**, *58*, 749-757.

(9) Rice, B. W.; Cable, M. D.; Nelson, M. B. *J. Biomed. Optics* **2001**, *6*, 432-440.

(10) Branchini, B. R.; Southworth, T. L.; Khattak, N. F.; Michelini, E.; Roda, A. *Anal. Biochem.* **2005**, *345*, 140-148.

(11) Branchini, B. R.; Ablamsky, D. M.; Rosenman, J. M.; Uzasci, L.; Southworth, T. L.; Zimmer, M. *Biochemistry* **2007**, *46*, 13847-13855.

(12) Nakatsu, T.; Ichiyama, S.; Hiratake, J.; Saldanha, A.; Kobashi, N.; Sakata, K.; Kato, H. *Nature* **2006**, *440*, 372-376.

(13) Nakatani, N.; Hasegawa, J.; Nakatsuji, H. *J. Am. Chem. Soc.* **2007**, *129*, 8756-8765.

(14) Nakatsuji, H.; Hirao, K. *J. Chem. Phys.* **1978**, *68*, 2053.

(15) Nakatsuji, H. *Chem. Phys. Lett.* **1978**, *59*, 362.

(16) Nakatsuji, H. *Chem. Phys. Lett.* **1979**, *67*, 329.

(17) Nakatsuji, H. *Chem. Phys. Lett.* **1979**, *67*, 334.

(18) Nakatsuji, H. *Computational Chemistry, Revies of Current Trends*: Leszczynski, J. Ed., World Scientific, Singapore, 1996, pp 62-124.

(19) Ehara, M.; Hasegawa, J.; Nakatsuji, H. *Theory and Applications of Computational Chemistry*: Dykstra, C., Frenking, G., Kim, K., Scuseria, G. Eds., Elsevier Science, New York, 2006.

(20) Hasegawa, J.; Nakatsuji, H. *Radiation Induced Molecular Phenomena in Nucleic Acid: A Comprehensive Theoretical and Experimental Analysis*: Shukla, M., Leszczynsk, J. Eds., Springer, 2008.

(21) Fujimoto, K.; Hasegawa, J.; Hayashi, S.; Kato, S.; Nakatsuji, H. *Chem. Phys. Lett.* **2005**, *414*, 239-242.

(22) Fujimoto, K.; Hasegawa, J.; Hayashi, S.; Nakatsuji, H. *Chem. Phys. Lett.* **2006**, *432*, 252-256.

(23) Fujimoto, K.; Hayashi, S.; Hasegawa, J.; Nakatsuji, H. *J. Chem. Theory Comput.* **2007**, *3*, 605-618.

(24) Fujimoto, K.; Hasegawa, J.; Nakatsuji, H. *Chem. Phys. Lett.* **2008**, *462*, 318-320.

(25) Das, A. K.; Hasegawa, J.; Miyahara, T.; Ehara, M.; Nakatsuji, H. *J. Comp. Chem.* **2003**, *24*, 1421.

(26) Hasegawa, J.; Fujimoto, K.; Swerts, B.; Miyahara, T.; Nakatsuji, H. *J. Comp. Chem.* **2007**, *28*, 2443-2452.

(27) Conti, E.; Franks, N. P.; Brick, P. *Structure* **1996**, *4*, 287-298.

(28) Branchini, B. R.; Magyar, R. A.; Murtiashaw, M. H.; Anderson, S. M.; Zimmer, M. *Biochemistry* **1998**, *37*, 15311-15319.

(29) Branchini, B. R.; Magyar, R. A.; Murtiashaw, M. H.; Anderson, S. M.; Helgerson, L. C.; Zimmer, M. *Biochemistry* **1999**, *38*, 13223-13230.

(30) Branchini, B. R.; Southworth, T. L.; Murtiashaw, M. H.; Boije, H.; Fleet, S. E. *Biochemistry* **2003**, *42*, 10429-10436.

(31) Branchini, B. R.; Southworth, T. L.; Murtiashaw, M. H.; Magyer, R. A.; Gonzalez, S. A.; Ruggiero, M. C.; Stroh, J. G. *Biochemistry* **2004**, *43*, 7255-7262.

(32) Branchini, B. R.; Magyar, R. A.; Murtiashaw, M. H.; Portier, N. C. *Biochemistry* **2001**, *40*, 2410-2418.

(33) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.

(34) Nakatsuji, H. *Chem. Phys.* **1983**, *75*, 425.

(35) Dunning, T. H.; Hay, P. J. *Modern Theoretical Chemistry*: Schaefer, H. F. Ed., Plenum, New York, 1976, p. 1-28.

(36) Franks, N. P.; Jenkins, A.; Conti, E.; Lieb, W. R.; Brick, P. *Biophys. J.* **1998**, *75*, 2205-2211.

(37) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.;

Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian03. Gaussian, Inc., Pittsburgh, 2003.

(38) Ponder, J. W.; Richards, F. M. *J. Comp. Chem.* **1987**, *8*, 1016-1024.

(39) Kundrot, C. E.; Ponder, J. W.; Richards, F. M. *J. Comp. Chem.* **1991**, *12*, 402-409.

(40) Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933-5947.

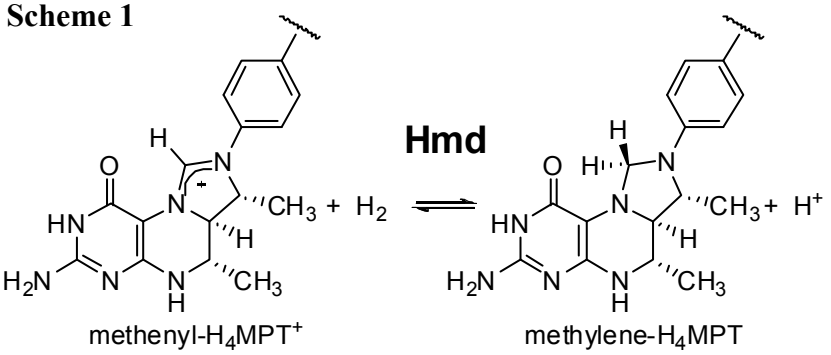
Chapter 4

Theoretical Study of the Iron Sulfur Cluster-Free Hydrogenase (Hmd): What is the Active Center of Hmd?

4.1 Introduction

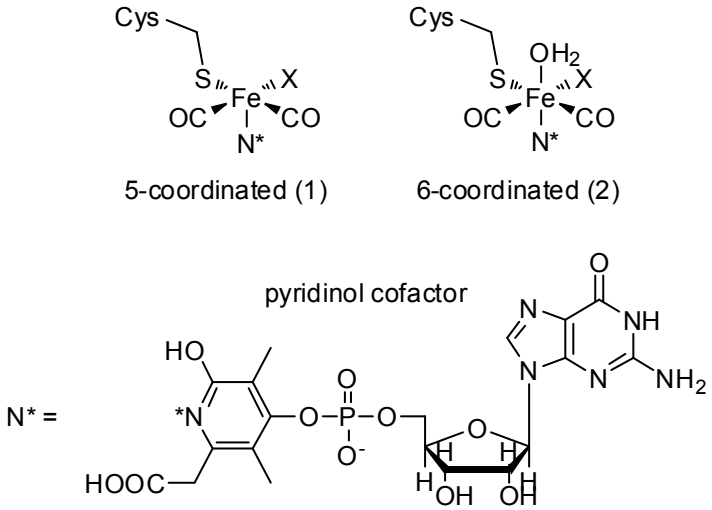
Iron-sulfur-cluster-free hydrogenase Hmd (H_2-forming methylenetetrahydromethanopterin dehydrogenase) catalyzes the reduction of N^5,N^{10}-methenyltetrahydromethanopterin (methenyl-H_4MPT$^+$) into N^5,N^{10}-methylene- tetrahydromethanopterin (methylene-H_4MPT), in which one H atom of dihydrogen molecule is introduced to methenyl-H_4MPT but the other becomes proton (Scheme 1).[1-3] Because this reaction involves reversible H_2 activation, it is of considerable importance that this is a biological fuel cell system. Considering that the

Scheme 1



iron-sulfur clusters are often found in biological red/ox systems to mediate electron transfer between active site and red/ox partner,[4] it is of considerable interest that Hmd does not need such iron-sulfur cluster to perform its catalytic reaction.[5,6]

Scheme 2



Recent crystallographic[5] and spectroscopic studies[6,7] have elucidated that the Hmd active site is either a five- or six-coordinate mononuclear iron complex which consists of two CO ligands, Cys176, pyridinol cofactor, and one unknown X ligand (Scheme 2). There are many important issues to be investigated in Hmd, as follows: (1) Mössbauer study suggested that the iron center is low-spin Fe(0) or low-spin Fe(II).[6a] Another IR study suggested that low-spin Fe(II) is a plausible candidate.[6b] But it has not been completely concluded which is involved in Hmd. (2) One ligand still remains unknown.[5] And (3) though monodentate pyridinol was proposed previously,[5a] bidentate acylpyridinol was recently proposed.[5b] It is not clear which is correct.

From kinetic studies, the reaction free energy ΔG^{o} of H_2/H^+ exchange is −5.5 kJ/mol and the activation barrier is 52 kJ/mol.[8] The H_2 activation reaction is inactivated by exposure of ultraviolet light.[9] This has been considered that the dissociation of one CO ligand from the Hmd active site is induced by the UV light exposure and it binds with the active site of another Hmd to function as concerted inhibitor. Recent IR study showed that the CO-inhibited form of Hmd contains three CO ligands in its active site.[7] The H_2 activation reaction is also inactivated by existence of cyanide anion (CN^-), which is bound to

the Hmd active site to function as non-concerted inhibitor.[9] Crystallographic study showed that the unknown X ligand is substituted by CN ligand in CN-inhibited form of Hmd.[5]

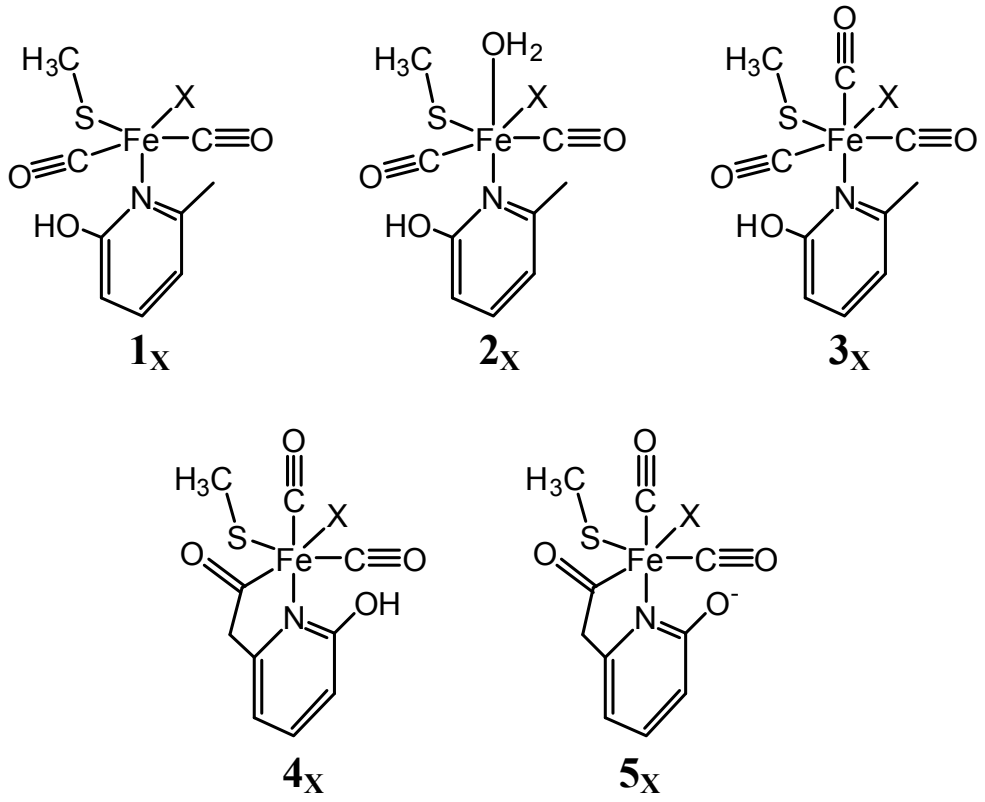
Because the stretching frequency of CO ligand reflects the geometry, the electronic structure of metal center, and the environment, the IR and raman spectra are very useful to investigate the Hmd active site. Actually, the IR spectra of Hmd have been recently proposed for the active form, the CO-inhibited form, and the CN-inhibited form.[7] Thus, it is worthy of theoretically investigating the CO stretching frequencies to clarify the geometry and the electronic structure of Hmd active site. In this study, we theoretically investigated the geometry, the electronic structure, and the CO stretching frequency of Hmd active site to determine what is the unknown X ligand and what iron complex is involved as the active site.

4.2 Computational Details

First, we constructed the possible models of Hmd active site, as shown in Scheme 3. In the monodentate pyridinol form, five-coordinate complex $\mathbf{1_X}$ and six-coordinate complex $\mathbf{2_X}$ are considered as possible candidates for the active form of Hmd (denoted as complex **A**, hereafter). Model complex having three CO ligands $\mathbf{3_X}$ is considered as a possible candidate for the CO-inhibited form of Hmd (denoted as complex $\mathbf{I_{CO}}$, hereafter). We employed hydroxide (OH^-), water (H_2O), and ammonia (NH_3) as plausible candidate for the unknown X ligand because the XENES study indicated that the Fe-X bond length is in the range of Fe-N and Fe-O bonds.[10] Cyanide complexes $\mathbf{1_{CN}}$ and $\mathbf{2_{CN}}$ are considered as possible candidates for the CN-inhibited form of Hmd (denoted as complex $\mathbf{I_{CN}}$, hereafter). In the bidentate pyridinol form, six-coordinate complex $\mathbf{4_X}$ is considered as a possible candidate for **A**, where X is OH^-, H_2O, and NH_3. The complex having three CO ligands $\mathbf{4_{CO}}$ and the cyanide complex $\mathbf{4_{CN}}$ are considered as possible candidates for $\mathbf{I_{CO}}$ and $\mathbf{I_{CN}}$, respectively. The

bidentate acyl-pyridone form $\mathbf{5_X}$ (X = OH^-, H_2O, and NH_3), $\mathbf{5_{CO}}$, and $\mathbf{5_{CN}}$ are also considered as possible candidates for **A**, $\mathbf{I_{CO}}$, and $\mathbf{I_{CN}}$, respectively.

Scheme 3. Computational models of Hmd active site



Next, we performed the geometry optimizations for each complexes with the DFT (B3LYP)[11-13] method. Basis sets we used were Stuttgart/Dresden (311111/22111/411/1) ECP10MDF for Fe,[14] 6-311G** for H, 6-311G* for C, and 6-311+G* for N, O, and S. The CO stretching frequency was calculated with B3LYP method and the same basis sets. In the energy calculations, effect of protein environment was evaluated by employing polarized continuum model (PCM)[15] with dielectric constant of 2.0 and 4.0. All calculations were performed by Gaussian 03 program package.[16]

4.3 Results and Discussions

4.3.1 Geometrical Feature and Oxidation State of Hmd Active Site

In Fe(0) complex, the pyridinol ligand dissociates from the iron center during geometry optimization to afford a planar four-coordinate structure; see Figure A1. This is consistent with our understanding that d^8 transition metal complexes with strong ligands tend to take a four-coordinate planar structure. The four-coordinate geometry is considerably different from the experimental structure in which at least five ligands apparently coordinate with the iron center.[5] In Fe(II) complex, on the other hand, the optimized geometry keeps five- or six-coordinate structure, which is consistent with the experimental structure. Thus, it should be clearly concluded that Hmd does not contain a low-spin Fe(0) center. We will concentrate on Fe(II) complexes hereafter.

We optimized various five- and six-coordinate Fe(II) complexes including monodentate pyridinol ($\mathbf{1_X}$, $\mathbf{2_X}$ and $\mathbf{3_X}$)[5a] and bidentate acyl-pyridinol ($\mathbf{4_X}$),[5b] and bidentate acyl-pyridone ($\mathbf{5_X}$),[5b] as displayed in Figure 1. We used subscripts OH, WT, AM, CN, and CO for OH^-, H_2O, NH_3, CN^-, and CO ligands, respectively, hereafter. Here, it is mentioned that the geometry optimization of $\mathbf{4_{OH}}$ leads to $\mathbf{5_{WT}}$ via proton transfer from the OH group of pyridinol to the OH ligand; in other words, $\mathbf{4_{OH}}$ is the same as $\mathbf{5_{WT}}$.

Complex $\mathbf{1_X}$ takes a distorted square pyramidal structure and complexes $\mathbf{2_X}$, $\mathbf{3_X}$, $\mathbf{4_X}$, and $\mathbf{5_X}$ take octahedral-like structures. Important bond lengths and bond angles of these complexes were summarized in Table A1 and A2. In $\mathbf{1_X}$, though the important bond lengths were consistent with the experimental values the important bond angles were much different from the experimental values. In $\mathbf{2_X}$, $\mathbf{3_X}$, $\mathbf{4_X}$, and $\mathbf{5_X}$, the important bond lengths and bond angles were both consistent with the experimental values. Thus, complex $\mathbf{1_X}$ was ruled out from the candidate of **A**, hereafter.

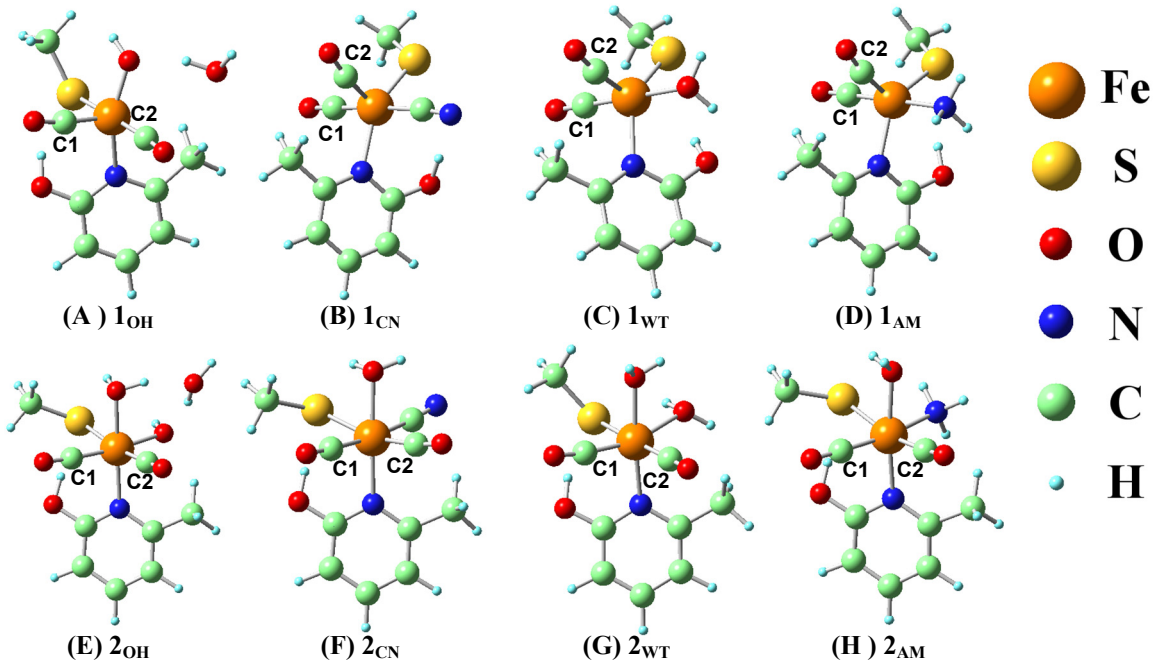


Figure 1. Optimized geometries of iron(II) complexes with DFT (B3LYP) method, (A) $\mathbf{1}_{OH}$, (B) $\mathbf{1}_{CN}$, (C) $\mathbf{1}_{WT}$, (D) $\mathbf{1}_{AM}$, (E) $\mathbf{2}_{OH}$, (F) $\mathbf{2}_{CN}$, (G) $\mathbf{2}_{WT}$ and (H) $\mathbf{2}_{AM}$.

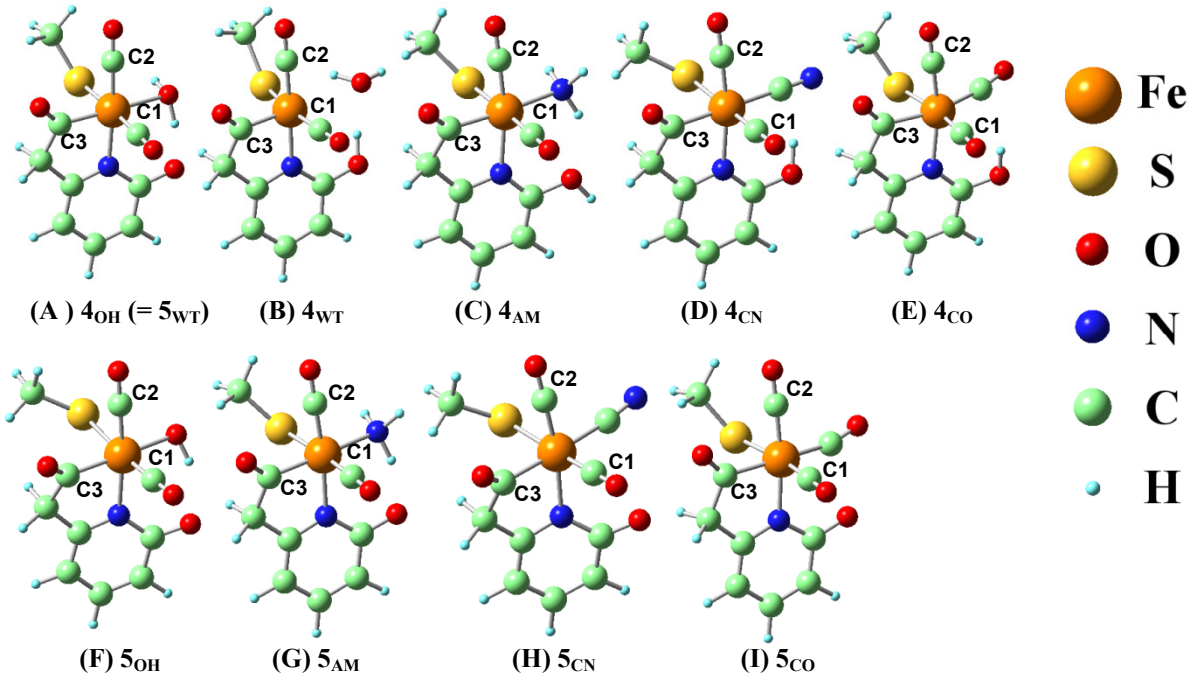


Figure 2. Optimized geometries of iron(II) complexes with DFT (B3LYP) method, (A) $\mathbf{4}_{OH}$ (= $\mathbf{5}_{WT}$), (B) $\mathbf{4}_{WT}$, (C) $\mathbf{4}_{AM}$, (D) $\mathbf{4}_{CN}$, (E) $\mathbf{4}_{CO}$, (F) $\mathbf{5}_{OH}$, (G) $\mathbf{5}_{AM}$, (H) $\mathbf{5}_{CN}$, and (I) $\mathbf{5}_{CO}$

4.3.2 Calculated CO Stretching Frequencies of Hmd Active Center

To investigate what is the unknown X ligand, we compare here the CO stretching frequency (ν_{CO}) between the calculated value of model complexes and the experimental value.[7] We calculated the ν_{CO} values of $\mathbf{2_X}$, $\mathbf{3_X}$, $\mathbf{4_X}$, and $\mathbf{5_X}$ with the DFT(B3LYP) method. When X is OH^-, we added one explicit water molecule, because OH^- ligand forms strong hydrogen bond with solvent water. Because the DFT-calculated ν_{CO} is in general larger than the experimental value, direct comparison is difficult. Here, we examined the relation between the DFT-calculated ν_{CO} and the experimental values, as shown in Figure 1.[17]

Previous experimental studies suggested that the CN^- ligand occupies the same position of unknown X ligand; in other words, $\mathbf{2_{CN}}$ and $\mathbf{2_X}$ are similar to $\mathbf{I_{CN}}$ and **A**, respectively.[5,9] Thus, it is likely that the relation of $\mathbf{2_X}$ vs. **A** is similar to that of $\mathbf{2_{CN}}$ vs. $\mathbf{I_{CN}}$, if **A** involves X as unknown ligand. As shown in Figure 1(A), the relation of $\mathbf{2_{OH}}$ vs. **A** is almost the same as that of $\mathbf{2_{CN}}$ vs. $\mathbf{I_{CN}}$, though the relations of $\mathbf{2_{WT}}$ vs. **A** and $\mathbf{2_{AM}}$ vs. **A** are much different from that of $\mathbf{2_{CN}}$ vs. $\mathbf{I_{CN}}$. If $\mathbf{2_X}$ is a good candidate for **A**, the relation of $\mathbf{3_X}$ vs. $\mathbf{I_{CO}}$ would be also similar to the relations of $\mathbf{2_{CN}}$ vs. $\mathbf{I_{CN}}$ and $\mathbf{2_X}$ vs. **A**. Apparently, the relation of $\mathbf{3_{OH}}$ vs. $\mathbf{I_{CO}}$ is almost the same as those of $\mathbf{2_{CN}}$ vs. $\mathbf{I_{CN}}$ and $\mathbf{2_{OH}}$ vs. **A** (Figure 1(B)). However, the relations of $\mathbf{3_{AM}}$ vs. $\mathbf{I_{CO}}$ and $\mathbf{3_{WT}}$ vs. $\mathbf{I_{CO}}$ are much different from that of $\mathbf{2_{CN}}$ vs. $\mathbf{I_{CN}}$. Though the Fe-OH_2 distance (2.09 Å) of $\mathbf{2_{OH}}$ is much shorter than the experimental value (2.7 Å),[5a] $\mathbf{2_{OH}}$ is a plausible candidate in the monodentate pyridinol form; the ν_{CO} values do not depend very much on the Fe-OH_2 distance (see Figure A1).

For the same reason as mentioned above, if **A** involves X as unknown ligand and the bidentate acyl-pyridinol complex $\mathbf{4_{CN}}$ is a good model of $\mathbf{I_{CN}}$, the relations of $\mathbf{4_X}$ vs **A** and $\mathbf{4_{CO}}$ vs. $\mathbf{I_{CO}}$ would be similar to that of $\mathbf{4_{CN}}$ vs. $\mathbf{I_{CN}}$. As shown in Figure 1(C), the relation of $\mathbf{4_{OH}}$ (= $\mathbf{5_{WT}}$) vs. **A** is almost the same as that of $\mathbf{4_{CN}}$ vs. $\mathbf{I_{CN}}$. However, the relations of $\mathbf{4_{WT}}$ vs. **A**

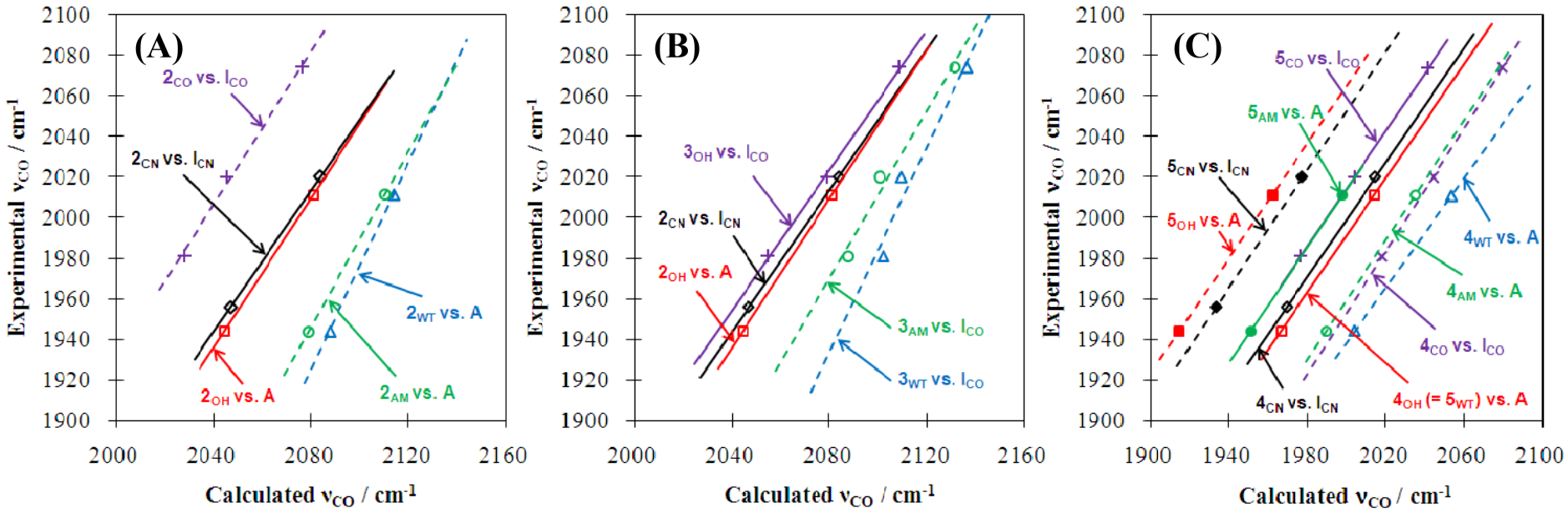


Figure 2. Calculated CO stretching frequencies ν_{CO} of various model compounds vs. experimental data,[7] (A) $\mathbf{2_X}$ vs. **A**, (B) $\mathbf{3_X}$ vs. $\mathbf{I_{CO}}$, and (C) $\mathbf{4_X}$ (or $\mathbf{5_X}$) vs. **A**. Solid and dashed lines indicate successful relation and unsuccessful one, respectively.

and $\mathbf{4_{AM}}$ vs. **A** are much different from that of $\mathbf{4_{CN}}$ vs. $\mathbf{I_{CN}}$. Also, the relation of $\mathbf{4_{CO}}$ vs. $\mathbf{I_{CO}}$ is somewhat different from that of $\mathbf{4_{CN}}$ vs. $\mathbf{I_{CN}}$, though the relation of $\mathbf{5_{CO}}$ vs. $\mathbf{I_{CO}}$ is almost the same as those of $\mathbf{4_{CN}}$ vs. $\mathbf{I_{CN}}$ and $\mathbf{4_{OH}}$ (= $\mathbf{5_{WT}}$) vs. **A**.[18] Similarly, the relation of $\mathbf{4_{AM}}$ vs. **A** is much different from $\mathbf{4_{CN}}$ vs. $\mathbf{I_{CN}}$, but that of $\mathbf{5_{AM}}$ vs. **A** is similar to that of $\mathbf{4_{CN}}$ vs. $\mathbf{I_{CN}}$. The relations of $\mathbf{5_{OH}}$ vs. **A** and $\mathbf{5_{CN}}$ vs. $\mathbf{I_{CN}}$ are much different from those of $\mathbf{5_{CO}}$ vs. $\mathbf{I_{CO}}$ and $\mathbf{4_{CO}}$ vs. $\mathbf{I_{CO}}$. From these results, it is concluded that $\mathbf{4_{OH}}$ (= $\mathbf{5_{WT}}$) and $\mathbf{5_{AM}}$ are also plausible candidates for **A**. However, ammonia is not often found as a ligand of metalloenzymes, to our knowledge. Thus, the water complex $\mathbf{5_{WT}}$ is the best candidate for the Hmd active complex in the acyl-pyridone form.

4.4 Conclusion

In this report, we theoretically investigated the geometry, the electronic structure, and the CO stretching frequency of Hmd active site. Though it has been experimentally proposed that the Hmd active site takes five- or six-coordinate iron center the low-spin Fe(0) complex takes planar-four coordinate structure. Thus, the low-spin Fe(0) complex was ruled out from the candidate of Hmd active site. The five-coordinate Fe(II) complex takes distorted square pyramidal structure. Though the important bond lengths were calculated to be consistent with the experimental structure the important bond angles were much different from the experimental structure. Thus, the five-coordinate Fe(II) complex was also ruled out from the candidate of Hmd active site.

The calculated CO stretching frequencies indicated that either the acyl-pyridone form and water or acyl-pyridinol form and hydroxide are the best candidates of Hmd active site and unknown X ligand, respectively. Also, mono-dentate pyridinol form and hydroxide are another plausible candidates of Hmd active site and unknown X ligand, respectively.

4.5 Appendix

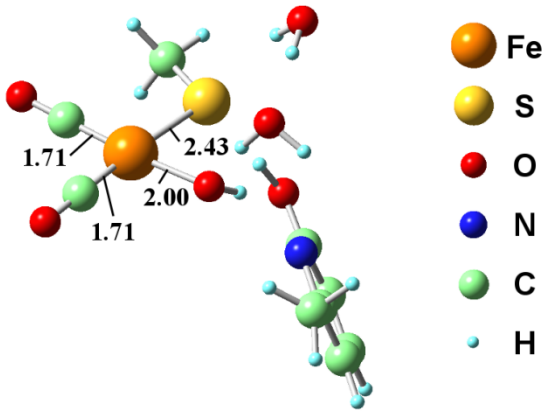


Figure A1. Optimized geometries of iron(0)—hydroxide complex with DFT (B3LYP) method.

Table A1. Important bond lengths and bond angles of optimized geometries of $\mathbf{1}_X$, $\mathbf{2}_X$ and $\mathbf{3}_X$. Bond angle A/B represents angle of A-Fe-B.

Species	Bond lengths / Å						Bond angles / degree											
	Fe-N	Fe-C1	Fe-C2	Fe-S	Fe-X	Fe-OH_2	C1/C2	C1/S	C2/X	S/X	N/C1	N/C2	N/S	N/X	H_2O/C1	H_2O/C2	H_2O/S	H_2O/X
$\mathbf{1}_{OH}$	2.10	1.76	1.87	2.29	1.88	-	91.6	91.4	87.1	95.9	103.3	91.1	84.6	157.3	-	-	-	-
$\mathbf{1}_{WT}$	2.08	1.81	1.83	2.16	2.09	-	94.8	91.9	88.0	85.1	94.2	102.1	136.9	86.9	-	-	-	-
$\mathbf{1}_{AM}$	2.09	1.82	1.81	2.17	2.08	-	93.6	92.4	90.4	84.7	92.3	104.4	135.1	87.7	-	-	-	-
$\mathbf{1}_{CN}$	2.12	1.83	1.77	2.19	1.95	-	96.4	92.7	88.6	85.9	90.7	105.5	149.7	88.1	-	-	-	-
$\mathbf{2}_{OH}$	2.17	1.84	1.85	2.40	1.96	2.09	90.4	90.7	92.7	85.6	93.3	92.2	96.4	91.4	88.7	87.9	83.5	86.6
$\mathbf{2}_{WT}$	2.12	1.84	1.88	2.35	2.09	2.10	91.9	89.8	98.1	80.1	95.0	89.8	90.4	97.7	89.8	89.8	89.8	77.5
$\mathbf{2}_{AM}$	2.15	1.84	1.87	2.40	2.07	2.11	90.9	89.6	97.6	81.3	91.9	93.1	100.5	90.9	94.8	85.8	80.7	82.7
$\mathbf{2}_{CN}$	2.17	1.86	1.83	2.43	1.95	2.12	91.9	90.7	92.0	84.2	92.4	94.3	99.3	92.7	90.8	86.7	79.6	84.0
Exptl.[6a]	2.1	1.8	1.8	2.4	1.8	2.7	94	79	114	74	102	91	86	94	72	91	90	90
Species	Bond lengths / Å						Bond angles / degree											
	Fe-N	Fe-C1	Fe-C2	Fe-S	Fe-X	Fe-C3	C1/C2	C1/S	C2/X	S/X	N/C1	N/C2	N/S	N/X	C3/C1	C3/C2	C3/S	C3/X
$\mathbf{3}_{OH}$	2.25	1.84	1.86	2.39	1.96	1.82	90.7	91.3	92.2	85.6	91.2	90.2	92.4	91.2	94.1	93.8	83.4	83.3
$\mathbf{3}_{WT}$	2.16	1.85	1.88	2.36	2.10	1.84	90.6	92.1	97.3	79.9	92.4	89.5	90.8	89.5	90.1	94.0	85.6	87.5
$\mathbf{3}_{AM}$	2.21	1.85	1.86	2.40	2.08	1.83	90.8	91.3	97.2	80.9	89.6	92.5	96.0	88.9	91.8	92.3	79.2	89.1

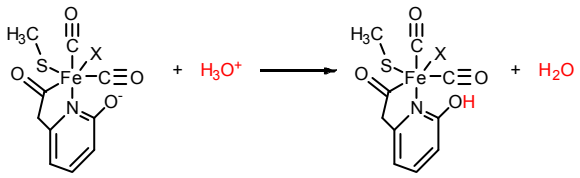
Table A2. Important bond lengths and bond angles of optimized geometries of $\mathbf{4}_X$ and $\mathbf{5}_X$. Bond angle A/B represents angle of A-Fe-B.

Species	Bond lengths / Å						Bond angles / degree											
	Fe-N	Fe-C1	Fe-C2	Fe-S	Fe-X	Fe-C3	C1/X	C1/C3	C3/S	S/X	N/C1	N/C3	N/S	N/X	C2/C1	C2/C3	C2/S	C2/X
$\mathbf{4}_{OH}$ ($\mathbf{5}_{WT}$)	2.08	1.78	1.78	2.39	2.17	1.93	95.2	91.6	88.6	84.4	89.2	84.5	87.3	92.0	93.6	92.7	89.8	90.4
$\mathbf{4}_{WT}$	2.11	1.80	1.80	2.36	2.23	1.92	96.1	92.2	90.3	81.5	90.7	84.0	86.1	96.3	93.2	91.3	90.3	87.8
$\mathbf{4}_{AM}$	2.15	1.79	1.78	2.38	2.17	1.94	96.5	91.1	88.8	83.5	90.3	82.8	87.2	96.4	94.9	89.9	87.6	90.1
$\mathbf{4}_{CN}$	2.15	1.78	1.78	2.41	2.00	1.97	92.7	91.0	89.3	87.1	90.9	82.6	90.0	95.6	95.8	91.4	83.3	89.9
$\mathbf{4}_{CO}$	2.10	1.80	1.80	2.40	1.89	1.98	99.9	91.2	83.9	85.1	90.1	82.6	86.2	97.5	94.4	88.1	88.5	90.7
$\mathbf{5}_{OH}$	2.14	1.76	1.76	2.48	2.02	1.96	91.5	93.8	89.4	85.4	90.1	82.0	93.8	94.9	95.2	94.1	81.1	88.5
$\mathbf{5}_{AM}$	2.11	1.78	1.77	2.41	2.13	1.95	96.8	91.0	86.9	85.2	88.4	83.7	89.0	92.3	94.9	91.9	87.5	91.6
$\mathbf{5}_{CN}$	2.11	1.77	1.77	2.48	2.02	1.98	93.1	92.6	87.9	86.5	88.6	80.9	93.5	97.5	95.7	92.2	82.2	88.9
$\mathbf{5}_{CO}$	2.03	1.78	1.79	2.42	1.92	1.98	98.1	90.6	82.9	88.2	89.1	83.2	87.7	92.9	94.9	91.1	87.7	92.1
Exptl.[6b]	2.05	1.77	1.77	2.34	2.05	1.88	97	97	74	91	92	84	87	95	89	91	91	89

Table A3. Proton affinities[a] of $\mathbf{5_{OH}}$, $\mathbf{5_{AM}}$, $\mathbf{5_{CN}}$ and $\mathbf{5_{CO}}$ calculated with PCM method at DFT(B3LYP) level, where dielectric constant 76.8 (water) was employed. A proton affinity (PE) of $\mathbf{5_X}$ is defined by PE = $E(\mathbf{5_X}+H_3O^+) - E(\mathbf{4_X}+H_2O)$.

	Proton affinity / kcal mol^{-1}
$\mathbf{5_{OH}}$	68.2
$\mathbf{5_{AM}}$	39.1
$\mathbf{5_{CN}}$	47.8
$\mathbf{5_{CO}}$	30.9

a)



$\mathbf{5_X}$ $\mathbf{4_X}$

References

(1) Zirngibl, C.; Hedderich, R.; Thauer, R. K. *FEBS Lett.* **1990**, *261*, 112-116.

(2) Schwörer, B.; Thauer, R. K. *Arch. Microbiol.* **1991**, *155*, 459-465.

(3) Ma, K.; Zirngibl, C.; Linder, D.; Stetter, K. O.; Thauer, R. K. *Arch. Microbiol.* **1991**, *156*, 43-48.

(4) Siegbahn, P. E. M.; Tye, J. W.; Hall, M. B. *Chem. Rev.* **2007**, *107*, 4414-4435.

(5) Shima, S.; Pilak, O.; Vogt, S.; Schick, M.; Stagni, M. S.; Meyer-Klaucke, W.; Warkentin, E.; Thauer, R. K.; Ermler, U. *Science* **2008**, *321*, 572-575.

(6) (a) Shima, S.; Lyon, E. J.; Thauer, R. K.; Mienert, B.; Bill, E. *J. Am. Chem. Soc.* **2005**, *127*, 10430-10435. (b) Wang, X.; Li, Z.; Zeng, X.; Luo, Q.; Evans, D. J.; Pickett, C. J. *Chem. Commun.* **2008**, 3555.

(7) Lyon, E. J.; Shima, S.; Boecher, R.; Thauer, R. K.; Grevels, F.-W.; Bill, E.; Roseboom, W.; Albracht, S. P. J. *J. Am. Chem. Soc.* **2004**, *126*, 14239-14248.

(8) Zirngibl, C.; Hedderich, R.; Thauer, R. K. *FEBS Lett.* **1990**, *261*, 112-116.

(9) Lyon, E. J.; Shima, S.; Buurman, G.; Chowdhuri, S.; Batschauer, A.; Steinbach, K.; Thauer, R. K. *Eur. J. Biochem.* **2004**, *271*, 195-204.

(10) Korbas, M.; Vogt, S.; Meyer-Klaucke, W.; Bill, E.; Lyon, E. J.; Thauer, R. K.; Shima, S. *J. Biol. Chem.* **2006**, *281*, 30804-30813.

(11) Lee, C.; Yang, W.; Parr, R.G. *Phis. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785-789.

(12) Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. *Chem. Phys. Lett.* **1989**, *157*, 200-206.

(13) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648-5652.

(14) #SDD

(15) Cancès, E.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032-3041.

(16) Frisch, M. J.; et al. *Gaussian 03*, Revision C.02; Gaussian Inc.; Wallingford, CT, 2004.

(17) Calculated ν_{CO} values were scaled by 0.969 which was determined here for hydrogenase model complexes.

(18) It is likely that the relations of $\mathbf{4_{CO}}$ vs. $\mathbf{I_{CO}}$ and $\mathbf{5_{CN}}$ vs. $\mathbf{I_{CN}}$ deviate from that of $\mathbf{4_{OH}}$ (=$\mathbf{5_{WT}}$) vs. **A** because the deprotonation of $\mathbf{4_{CO}}$ easily occurs to produce $\mathbf{5_{CO}}$ and the protonation of $\mathbf{5_{CN}}$ easily occurs to form $\mathbf{4_{CN}}$ (see Table A3).

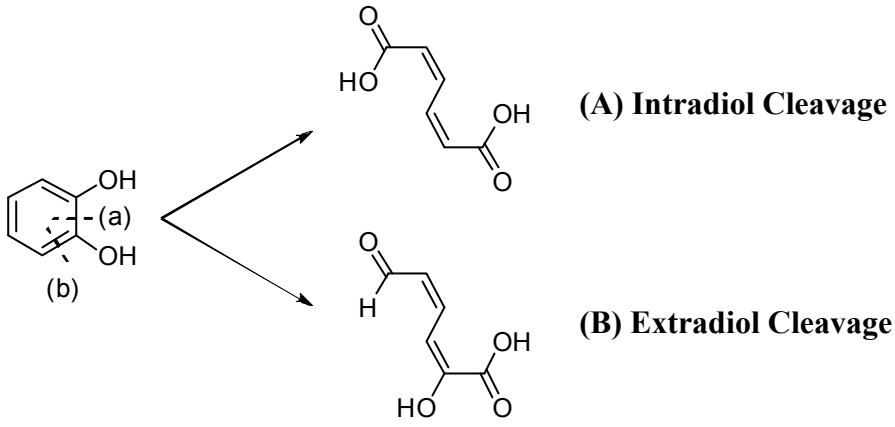
Chapter 5

Theoretical Study of Dioxygen Binding Process in Iron(III) Catechol Dioxygenase: "Oxygen Activation" vs. "Substrate Activation"

5.1 Introduction

Catechol dioxygenases, found in diverse range of soil bacteria, activate dioxygen molecule and catalyze the aromatic bond cleavage of catechol (1,2-dihydroxylbenzene) and its derivatives.[1-5] Their oxygenation catalyses are of considerable interest, as follows: The aromatic bond cleavage is in general difficult because of its considerably large stability due to the resonance structure. Also, dioxygen molecule is not reactive to most of organic compounds, because dioxygen molecule has triplet ground state but most of organic molecules have closed-shell singlet ground state. In spite of these difficulties, catechol dioxygenases successfully catalyze oxygenation reactions including the aromatic bond cleavage under mild reaction conditions. Catechol dioxygenases possess a non-heme iron complex in the active center, which plays important roles in the oxygenation catalysis. They are divided into two types with respect to the oxidation state of the iron center; the catechol dioxygenase with iron(III) center catalyzes 1,2-cleavage of catechol (intra-diol cleavage),[6] as shown in Scheme 1(A), and the other with iron(II) center catalyzes 2,3-cleavage (extra-diol cleavage),[7] as shown in Scheme 1(B). Though the dioxygen activation by iron(II) system is well established in heme proteins, heme protein which contains the iron(III) center is not reactive to the dioxygen molecule, as well known. In this regard, the activation mechanism of dioxygen molecule by the iron(III) catechol dioxygenase is of considerable interest.

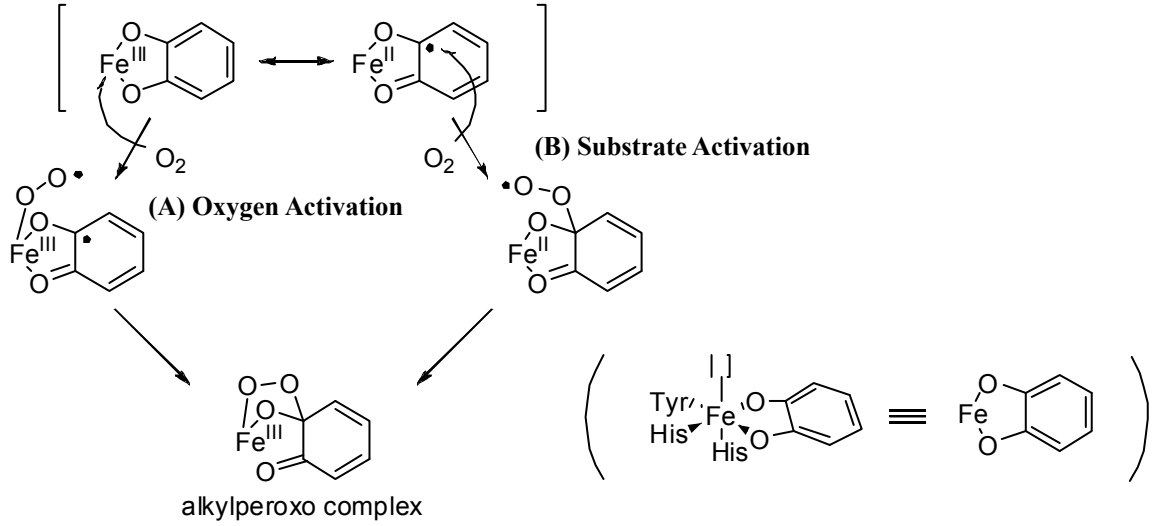
Scheme 1. Catalytic reaction of catechol dioxygenase



In the iron(III) catechol dioxygenase, the active site contains mononuclear iron(III) center, two histidine residues (His460 and His462), two tyrosine residues (Tyr408 and Tyr447), and water molecule.[8,9] The substrate catechol coordinates to the iron center with concomitant dissociation of Tyr447 and water molecule, to form distorted square pyramidal structure.[10] Then, the dioxygen molecule approaches the active center and the oxygenation reaction of catechol occurs to produce *cis*,*cis*-muconic acid. Interestingly, it is known that the iron center keeps high-spin character during the catalytic cycle.[11,12]

Oxygenation reaction and its catalytic mechanism of intra-diol cleavage have been well established for decades.[13-18] In the O_2 binding step, two different reaction courses were proposed; one is called "oxygen activation"[15,19-21] and the other is called "substrate activation".[16-18] In the "oxygen activation" mechanism, dioxygen molecule attacks the iron(III) center to form an iron-dioxygen complex, in which the charge transfer occurs from the catecholate moiety to the dioxygen moiety to afford iron(III)-superoxo complex, as shown in Scheme 2(A). In the "substrate activation" mechanism, the charge transfer occurs from the catecholate moiety to the iron(III) center to produce the iron(II)-semiquinonate radical,[18] which reacts with the dioxygen molecule to form a C-O bond between dioxygen molecule and

Scheme 2. Dioxygen activation mechanism previously proposed[15-21]



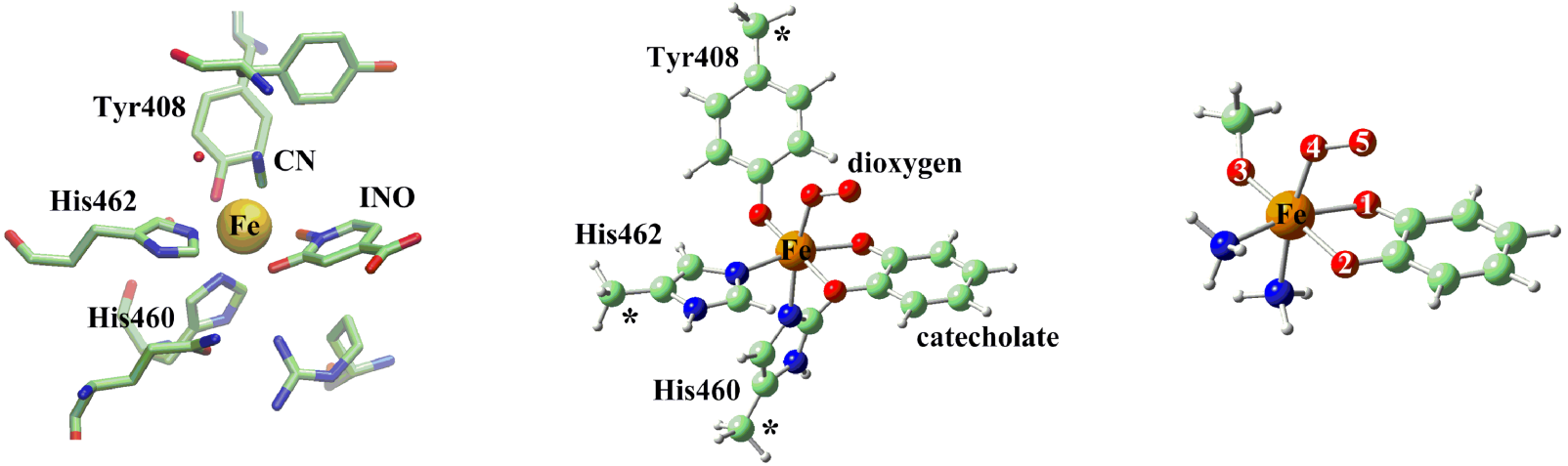
catecholate carbon atom, as shown in Scheme 2(B). From kinetic studies, the O_2 binding process rapidly occurs with the rate constant of 5×10^5 M^{-1} s^{-1} in the native protocatechuate-3,4-dioxygenase (3,4-PCD) from *pseudomonas putida*.[22] From recent spectroscopic studies, the iron-catecholate complexes are in equilibrium between the low-spin and high-spin states at low temperature. Interestingly, it was experimentally suggested that the low-spin state prefers the iron(II)-semiquinonate form but the high-spin state prefers the iron(III)-catecholate form.[19] Also, the O_2 binding process is reversible,[23] which is considered to be consistent with the formation of iron(III)-superoxo complex in the "oxygen activation" mechanism. Funabiki and co-workers[24,25] theoretically investigated the activation process with the UHF method and suggested that the "oxygen activation" mechanism is likely. Siegbahn and co-workers[26] theoretically investigated the whole catalytic cycle of intra-diol cleavage with the DFT method, where the dioxygen binding with iron(III) center was proposed in the intermediate. They also reported that the dioxygen attack to the catecholate carbon needed much larger activation barrier. On the other hand, the MCD spectroscopic and another DFT theoretical studies[27] suggested that the iron center

has the intermediate spin state (S = 3/2) through the anti-ferromagnetic coupling between the dioxygen molecule and the iron center, which induced spin-forbidden attack of O_2 on the catechol carbon atom. These suggestions support the "substrate activation" mechanism. Thus, the final conclusion has not been presented yet about the activation mechanism.

The detailed investigation about such approaching process provides us direct answer about the activation mechanism. In previous theoretical studies, however, the approaching process of dioxygen molecule to the iron(III) catecholate complex has not been investigated well. Moreover, multi-reference post-Hartree-Fock method, such as CASSCF/CASPT2 method, has not been employed to investigate the activation process of dioxygen molecule, though multi-configurational character must be taken into consideration of the interaction between iron and dioxygen molecule.

In this paper, we theoretically investigated the dioxygen activation process in the iron(III) catechol dioxygenase. We employed here complete-active-space-self-consistent-field (CASSCF) method[28,29] and its second-order perturbation theory[30,31] (CASPT2) to directly evaluate the multi-configurational character of the Fe-O_2 interaction. Our main purposes are to present clear conclusion about the activation mechanism of dioxygen molecule, to clarify how dioxygen molecule is activated by the iron(III) center, and what roles the iron(III) center and the catecholate substrate play in the activation process.

Scheme 3. (A) Experimental structure[10] of the active site in iron(III) catechol dioxygenase, (B) large model (LM), and (C) small model (SM)



(A) Experimental structure[a] **(B) Large model (LM)**[b] **(C) Small model (SM)**

[a] INO represents 2-hydroxyisonicotinic acid N-oxide. CN and INO are the inactive substrates mimicking dioxygen molecule and protocatechuate acid, respectively.

[b] In LM, the asterisk represents the Cβ atoms whose positions were fixed to those of the experimental structure.

5.2 Models and Computational Details

We employed here two computational models based on the experimental structure of protocatechuate-3,4-dioxygenase from *Pseudomonas putida* coordinated with cyanide (PDB-ID = 3PCL).[10] As shown in Scheme 3(A), the active site consists of Fe, 2-hydroxyisonicotinic acid *N*-oxide (INO), CN, His460, His462, and Tyr408.[32,33] In this structure, the INO and CN are included as the inhibitors instead of the 3,4-protocatechuate (3,4-dihydroxybenzoate) and dioxygen molecule, respectively. Considering large computational costs of the CASSCF and CASPT2 calculations, we employed two models, "large model (**LM**)" and "small model (**SM**)", where INO and CN were substituted for catecholate and dioxygen molecule, respectively, as shown in Scheme 3(B). In **LM**, amino-chains of His460, His462, and Tyr408 were removed, where the positions of the C_β atoms of His460, His462, and Tyr408 were fixed to those of the experimental structure and the C_β atoms were capped by hydrogen atoms. The direction of the C_β-H bond was taken to be the same as the C_α-C_β bond of the experimental structure and its C_β-H distance was optimized. In **SM**, His462 and His460 were substituted for ammonia (NH_3) and Tyr408 was substituted for methoxide (OCH_3), as shown in Scheme 3(C). The geometry of **SM** was taken to be the same as that of the corresponding moiety of **LM**. We employed **LM** in geometry optimization with the DFT method and **SM** in energy evaluation with the CASSCF and CASPT2 methods.

The geometry of **LM** was optimized by unrestricted density functional theory (UDFT) with B3LYP[34-36] functional in each spin state. We used (311111/22111/411/1) basis set for Fe with effective core potential (ECP) for 10 core electrons[37] and 6-31G* basis sets for C, N, O, and H, where diffuse functions of [1s1p] were added to O. This basis set system is called BAS-I hereafter. Better basis sets were used in energy evaluation; 6-311G* basis sets were

used for C, N, O, and H, where diffuse functions of [1s1p] were added to O, while the same basis set as that of BAS-I was used for Fe. This basis set system is called BAS-II.

In CASSCF calculation, we employed the active space with nine electrons in eight orbitals such as five d orbitals of Fe (d_{Fe}), π orbital of catecholate (π_{CAT}), and two π* orbitals of O_2 (π_p* and π_v*), where the subscripts p and v represent the π* orbital in the Fe-O-O plane and that vertical to the Fe-O-O plane, respectively. This CASSCF calculation is called CASSCF(9,8) hereafter. In energy evaluation, state-specific (SS) CASSCF/CASPT2 and state-averaged (SA) CASSCF/CASPT2 methods were employed, where the CASSCF wave function was employed to evaluate Mulliken charge and Mulliken spin population. We preliminarily carried out several SA-CASSCF calculations to check whether the potential energy curves are well calculated or not (see Figure A1 in appendix). In those calculations, we found that 3-state-averaged (3SA) CASSCF(9,8) calculation presents reliable potential energy curves for the O_2 binding process in the octet and quartet spin states. In the sextet spin state, 7SA-CASSCF(9,8) calculation presents reliable potential energy curve (see Figure A2). Thus, we applied the 3SA-CASSCF(9,8)/CASPT2 method to the octet and quartet spin states, and the 7SA-CASSCF(9,8)/CASPT2 method to the sextet spin state. CASSCF(9,9), CASSCF(11,9), CASSCF(11,10), and CASSCF(11,11) calculations were carried out to check whether the active space of 9 electrons in 8 orbitals presents correct results or not. CASSCF(9,8)/CASPT2-computed energy changes are similar to CASSCF(11,11)/CASPT2-computed ones in the octet spin state (see Figure A3), indicating that the active space of 9 electrons in 8 orbitals is proper choice and CASSCF(9,8)/CASPT2-computed results are reliable.

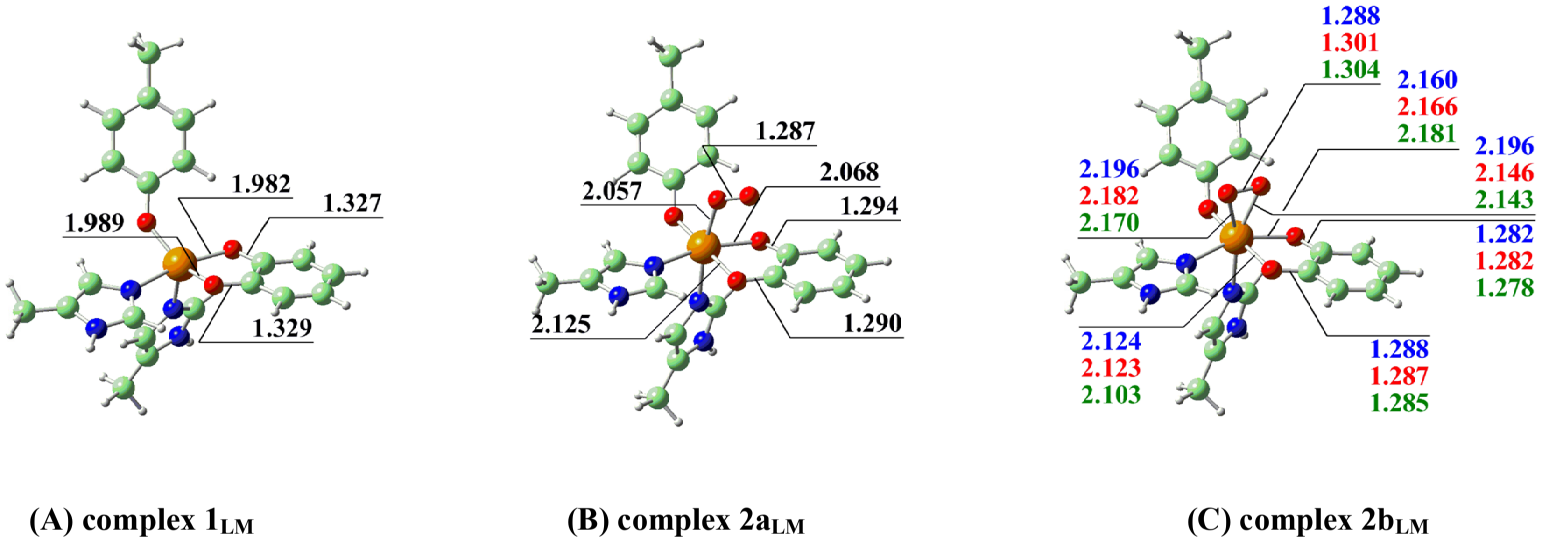
Because of the presence of intruder states, the level shift technique[38,39] was employed in the CASPT2 calculation. The shift value was taken to be 0.20 E_h when the SS-CASSCF

wave function was taken as a reference and 0.30 E_h when the SA-CASSCF wave function was taken as a reference. We examined several shift values and found that the values of 0.20 E_h and 0.30 E_h were the lowest which completely excluded the intruder states in SS-CASSCF(9,8)/CASPT2 and SA-CASSCF(9,8)/CASPT2 calculations, respectively (see Figure A4 and Figure A5).

We used *Gaussian03* program package[40] for DFT calculation and *Molcas* (version 6.4) program package[41] for CASSCF and CASPT2 calculations.

5.3 Results and Discussions

The optimized structures of iron(III) catecholate complex $\mathbf{1_{LM}}$ and iron(III) catecholate dioxygen complex $\mathbf{2_{LM}}$ are shown in Figure 1. In $\mathbf{2_{LM}}$, we found two types of equilibrium structures; one involves η^1-end-on coordinated dioxygen moiety ($\mathbf{2a_{LM}}$), which is the same as the previously proposed structure,[26] and the other involves η^2-side-on coordinated dioxygen moiety ($\mathbf{2b_{LM}}$), which has not been proposed both experimentally and theoretically in the iron(III) complex. We tried to optimize iron(II) semiquinonate dioxygen intermediate which was proposed in the substrate activation mechanism, but we failed it, as will be discussed below in detail. Geometries, electronic structures, and spin states of these species will be discussed below.



(A) complex 1_{LM} **(B) complex $2a_{LM}$** **(C) complex $2b_{LM}$**

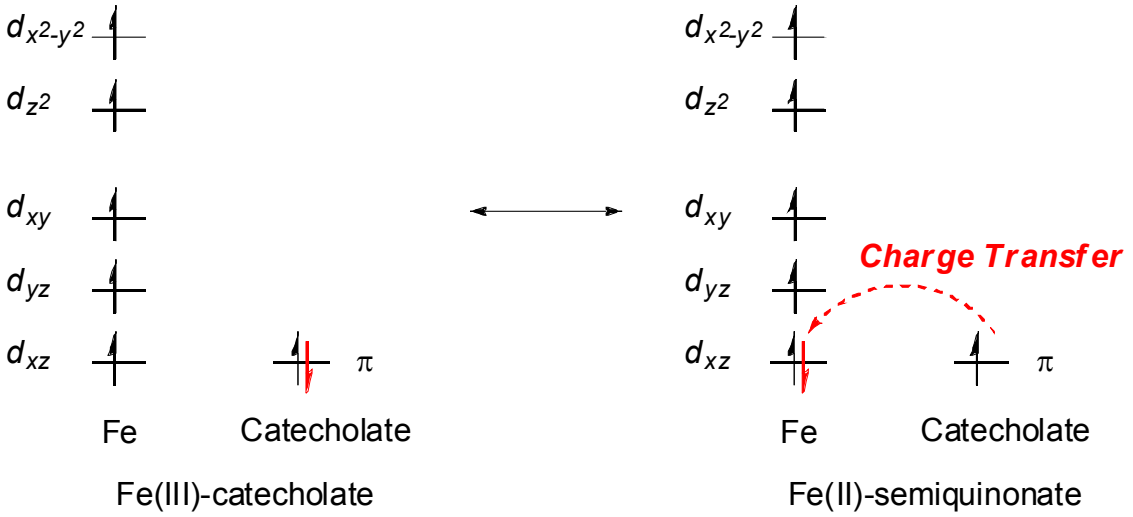
Figure 1. Optimized structures and important bond lengths[a)] of (A) iron(III)-catecholate complex [Fe(CAT)(MeIm)$_2$(MePhO)], $\mathbf{1_{LM}}$ in sextet spin state, (B) end-on type iron(III)-superoxo complex [Fe(CAT)(MeIm)$_2$(MePhO)(O$_2$)], $\mathbf{2a_{LM}}$ in sextet spin state, and (C) side-on type iron(III)-superoxo complex, $\mathbf{2b_{LM}}$ in octet (blue), sextet (red), and quartet (green) spin states, where CAT, MeIm, and MePhO represent catecholate, 4-methyl imidazol, and 4-methyl phenoxide, respectively, which are the models of protocatechuate, histidine, and tyrosine, respectively.

5.3.1 Geometry, Electronic Structure, and Spin State of Iron(III) Catecholate Complex 1.

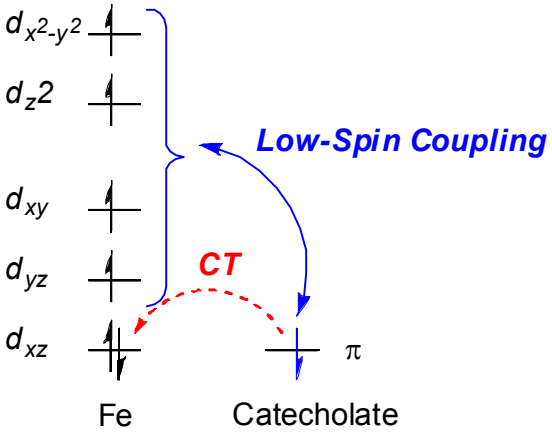
The equilibrium structure of the iron(III) catecholate complex $\mathbf{1}_{\mathbf{LM}}$ in the sextet spin state is optimized with the DFT/BAS-I method, as shown in Figure 1(A). This sextet spin state originates from the high spin iron(III) center. The Fe-O1 and Fe-O2 distances are 1.98 and 1.99 Å, respectively, which shows that both O atoms of catecholate strongly coordinate with the iron center. The O3 atom of Tyr408 model takes a position deviating from the xy-plane toward the z-direction. Thus, the geometry is understood to be distorted square pyramidal. The equilibrium structures of $\mathbf{1}_{\mathbf{LM}}$ in the quartet and doublet spin states are also optimized with the DFT/BAS-I method. The quartet and doublet spin states are calculated to be less stable than the sextet spin state by 1.5 and 22.2 kcal/mol, respectively, with the DFT/BAS-II method. Interestingly, the energy difference between the sextet and the quartet spin states is surprisingly small. It is noted that though the S^2 value (8.77) of the sextet spin state is similar to the exact value, the S^2 values (4.71 and 1.67) of the quartet and doublet spin states considerably deviate from their exact values, indicating that the spin contamination is small in the sextet spin state but large in the quartet and doublet spin states. Because the sextet spin state is the ground state in both DFT and CASPT2 calculations (see below), the spin contamination does not affect very much the equilibrium geometry of **1**.

The SS-CASSCF(7,6)/CASPT2 calculation of $\mathbf{1}_{\mathbf{SM}}$ was carried out to investigate the energy, where the active space consists of seven electrons in five d orbitals of Fe and π orbital of catecholate. The ground state is the sextet spin state, while the quartet and doublet spin states are calculated to be much more unstable than the sextet spin state by 10.4 and 46.8 kcal/mol, respectively. In the sextet spin state of $\mathbf{1}_{\mathbf{SM}}$, Mulliken spin population and Mulliken charge are 4.78e and 1.06e, respectively, for the iron center, and 0.11e and -0.85e,

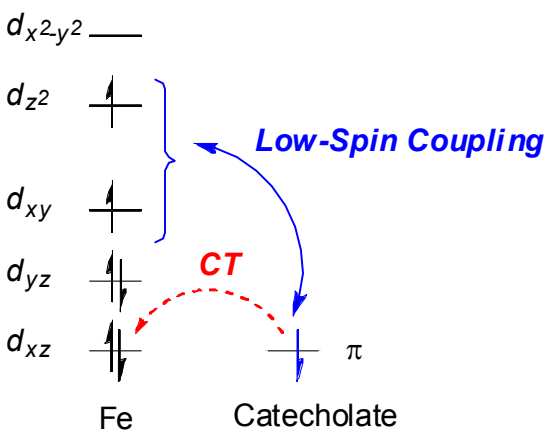
Scheme 4. Electronic structures of complex **1** in the (A) sextet, (B) quartet, and (C) doublet spin states



(A) sextet spin state



(B) quartet spin state



(C) doublet spin state

respectively, for the catecholate moiety, where CASSCF wave function was employed for evaluations of Mulliken spin population and Mulliken charge, hereafter. In the catecholate moiety, Mulliken spin population and Mulliken charge are 0.01e and 0.29e, respectively, for the aromatic ring, and 0.10e and -1.13e, respectively, for the phenol oxygens. It is noted that

the spin delocalization considerably occurs on the phenol oxygens but much less on the aromatic ring. From these results, it is clearly concluded that the iron center has high spin d^5 character, and radical character is little observed in the aromatic ring of the catecholate moiety (see Scheme 4(A)); in other words, **1** does not take the iron(II)-semiquinonate radical form in the sextet spin state. This is consistent with recent spectroscopic result that the iron(II)-semiquinonate-radical is not formed well in high-spin iron complex.[19] Consistent with this understanding, Mulliken charge (-0.85e) of the catecholate moiety is much more negative than that of the quartet spin state and the d orbital population of the iron center (5.4e) is much smaller than that of the quartet spin state in which the iron(II)-semiquinonate character is involved, as will be discussed below. In the quartet spin state, Mulliken spin populations are 3.53e and -0.58e and Mulliken charges are 1.12e and -0.54e for the iron center and the catecholate moiety, respectively. It is noted that though Mulliken charge of the iron center is similar to that of the sextet spin state the d orbital population (6.1e) is much larger than that (5.4e) in the sextet spin state. These results indicate that the iron center has high spin iron(II) character and one β-spin is localized in the catechol ring in the quartet spin state (see Scheme 4(B)). Though this character corresponds to the iron(II)-semiquinonate species, the quartet spin state is much less stable than the sextet spin state, as described above. In the doublet spin state, Mulliken spin populations are 1.31e and -0.32e and Mulliken charges are 1.07e and -0.44e for the iron center and the catecholate moiety, respectively. However, the doublet spin state is much more unstable than the sextet an quartet spin states. From these results, it should be concluded that the iron(II)-semiquinonate character is little involved in the ground state (sextet spin state) of iron(III)-catecholate complex but considerably in the quartet and doublet spin states. These results are also consistent with experimental result that the low-spin state prefers the iron(II)-semiquinonate radical state.[19]

5.3.2 Iron(III) Catecholate η^1-End-on Dioxygen Complex 2a.

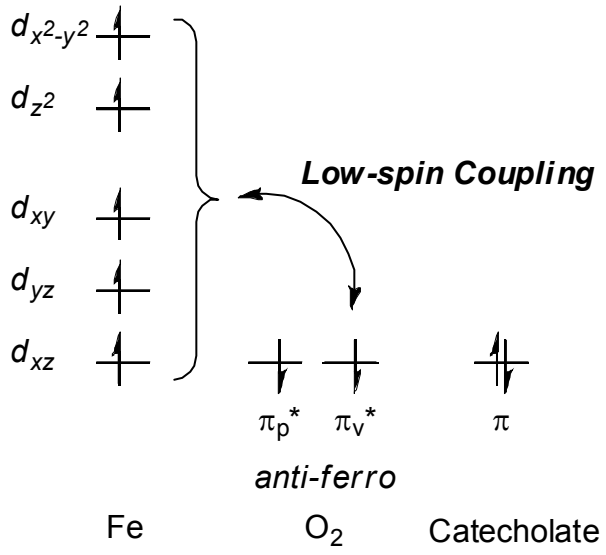
The optimized structure of the iron(III) catecholate dioxygen complex $\mathbf{2a_{LM}}$ in the sextet spin state is shown in Figure 1(B). In this geometry, the O4-O5 distance of the dioxygen moiety (1.29 Å) is much longer than that (1.208 Å) of free dioxygen molecule[42] but close to that (1.3 Å) of usual superoxide.[43] The Fe-O4 distance is 2.06 Å and the Fe-O5 distance is 2.99 Å. These geometrical features clearly show that $\mathbf{2a_{LM}}$ is typical η^1-end-on superoxo complex.

The geometry of $\mathbf{2a_{LM}}$ in the octet and quartet spin states could not be optimized with the DFT method, because their potential energy curves calculated with the DFT method are slightly repulsive along the Fe-O_2 distance in these spin states, as shown in Figure A6. Also, there is no energy minimum in these spin states. We investigated here the energy and electronic structure of **2a** in the octet and quartet spin states, using the geometry of the sextet spin state. According to the DFT/BAS-II calculation of $\mathbf{2a_{LM}}$, the octet and quartet spin states are more stable than the sextet spin state by 2.2 and 0.2 kcal/mol, respectively. Seemingly, these results are not consistent with the O_2 binding, because these results indicated that the ground state is the octet spin state but the potential energy curve is repulsive.

Because the DFT-computed results show the very small energy difference between these spin states, we carried out SS-CASSCF(9,8)/CASPT2 calculations of $\mathbf{2a_{SM}}$ to present more detailed results about spin state.

The SS-CASSCF(9,8)/CASPT2-computations indicate that the ground state is the sextet spin state, and the octet and quartet spin states are above the sextet spin state by 3.3 and 2.8 kcal/mol, respectively, unlike the DFT-computed results. The SS-CASSCF(9,8)-computed Mulliken spin populations are 4.80e, 1.04e, and 1.05e for the iron center, catecholate, and dioxygen moieties, respectively, in the octet spin state, 4.43e, 1.03e, and -0.56e, respectively,

Scheme 5. Low-spin coupling between the Fe center and the dioxygen molecule

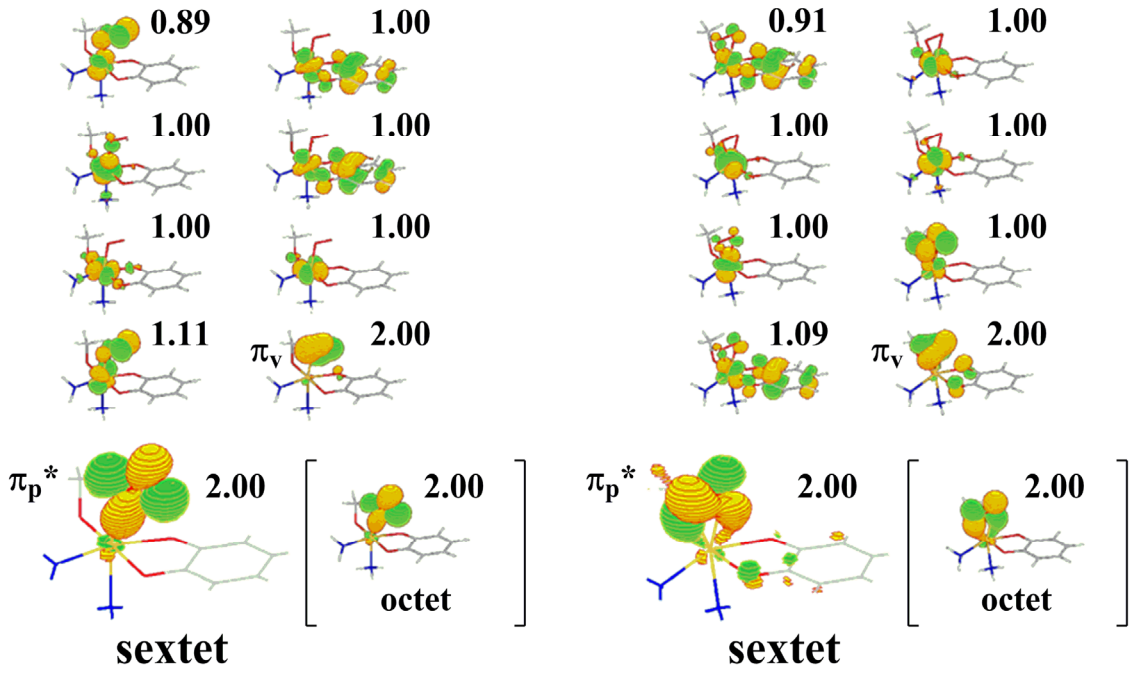


in the sextet spin state, and 4.01e, -0.55e, and -0.56e, respectively, in the quartet spin state. The spin population of the iron center moderately decreases upon going to the quartet spin state from the octet spin state. If the iron(III) center has intermediate or low spin state, the formal spin population should be 3.0 or 1.0, respectively. The present spin population on the iron center is much larger than these formal values of the intermediate and low spin states. These results clearly indicate that the sextet and quartet spin states of $\mathbf{2a_{SM}}$ arise from the low spin coupling between the sextet spin state of the iron(III) center and the triplet spin state of dioxygen molecule, as shown in Scheme 5. In all these spin states, the spin population of the catecholate moiety is larger in $\mathbf{2a_{SM}}$ than in $\mathbf{1_{SM}}$, while that of the dioxygen moiety in $\mathbf{2a_{SM}}$ is much lower than 2.0e which is the formal spin population of the triplet spin state. In $\mathbf{2a_{SM}}$, Mulliken charges are 1.18e, -0.27e, and -0.52e for the iron center, catecholate, and dioxygen moieties, respectively, in all spin states. It is noted that though the Mulliken charge of the iron center is little different between $\mathbf{1_{SM}}$ and $\mathbf{2a_{SM}}$ the Mulliken charge of the catecholate moiety is considerably less negative in $\mathbf{2a_{SM}}$ than in $\mathbf{1_{SM}}$ and Mulliken charge of dioxygen

moiety is considerably negative. These results indicate that the charge transfer occurs from the catecholate moiety to the dioxygen moiety to afford an iron(III)-superoxo complex. Mulliken spin populations of the aromatic ring and phenol oxygens of the catecholate moiety are 0.77e and 0.27e, respectively, in the octet spin state, 0.77e and 0.26e, respectively, in the sextet spin state, and -0.46e and -0.09e, respectively, in the quartet spin state. Mulliken charges of the aromatic ring and phenol oxygens are 0.69e and -0.96e, respectively, in all spin states. These results indicate that the spin population is largely localized in the aromatic ring, which is considered as the semiquinonate radical character. These results lead to clear conclusion that the semiquinonate radical character is little involved before O_2 binding but considerably after O_2 binding.

The natural orbitals in the sextet spin state, calculated with SS-CASSCF(9,8) method, are shown in Figure 2(A) with their occupation numbers.[44] These occupation numbers are similar to those of the octet and quartet spin states. The doubly occupied natural orbital mainly consists of the π_p* orbital of O_2, into which the d_{xz} orbital of Fe moderately mixes in a bonding way. This is the reasonable bonding interaction between the iron center and the dioxygen moiety. Besides this orbital, five d orbitals of Fe, π orbital of catecholate, and π_v* orbital of O_2 are found to be singly occupied, indicating that the sextet spin state involves the high spin d^5 character of the iron center, superoxo character of the dioxygen moiety, and the semiquinonate radical character of the catecholate moiety.

We tried to optimize another structure in which the dioxygen molecule is directly bound with the carbon atom of catecholate. Such species was proposed as the intermediate in the "substrate activation" mechanism.[16-18] However, we failed it. This is because the semiquinonate radical character is little involved in the iron(III)-catecholate complex **1**, as described above.



(A) η^1-end-on Fe-O_2 complex 2a **(B) η^2-side-on Fe-O_2 complex 2b**

Figure 2. Natural orbitals (NOs) calculated by the SS-CASSCF(9,8) method for (A) the η^1-end-on type iron(III)-superoxo complex $[Fe(CAT)(NH_3)_2(MeO)(O_2)]$ $\mathbf{2a_{SM}}$ and (B) the η^2-side-on type iron(III)-superoxo complex $\mathbf{2b_{SM}}$ in the sextet spin state

All these results support the "oxygen activation" mechanism but do not the "substrate activation" mechanism.

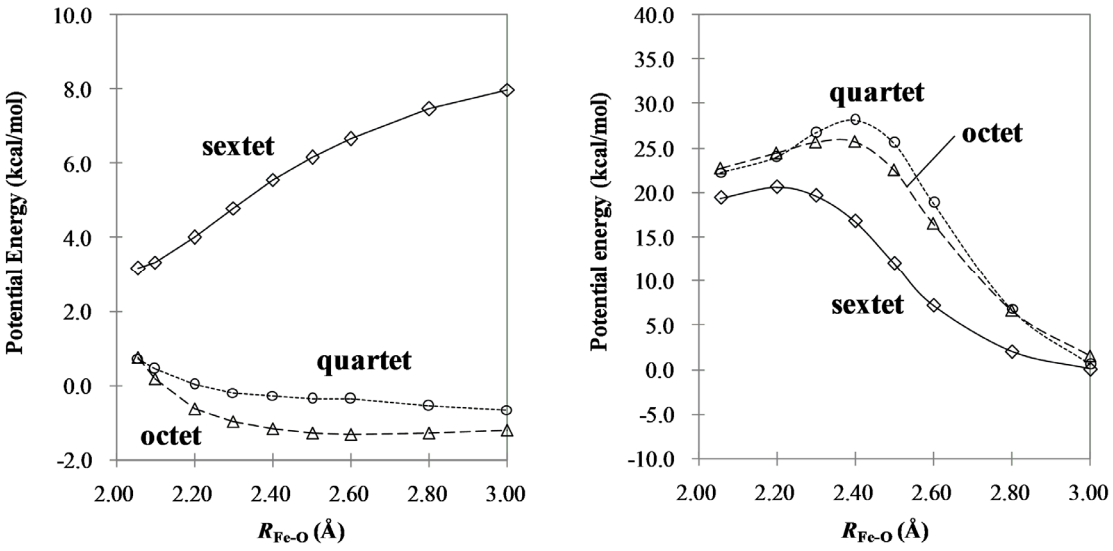
5.3.3 Potential Energy Curve of Dioxygen Binding Process.

It is worthy investigating potential energy curve (PEC) of O_2 binding process starting from **1** to **2a**. This PEC provides clear answer about the activation mechanism of dioxygen molecule and also well understanding of the activation process how dioxygen molecule is activated by the iron(III) complex. We performed the partial geometry optimization of **LM** with the DFT method, employing the Fe-O distance as the reaction coordinate in the octet, sextet, and quartet spin states. The PECs of **SM** were evaluated with the B3LYP/BAS-II

method in all spin states, as shown in Figure 3(A). The DFT-calculated O_2 binding process is slightly endothermic in the octet and quartet spin states, while it is slightly exothermic in the sextet spin state. The iron(III)-catecholate-dioxygen adduct has sextet spin state, as reported recently[26] and shown above by our SS-CASSCF/CASPT2 calculation. On the other hand, the long separated system consisting of the iron(III) catecholate complex and dioxygen molecule has octet spin state: remember that the iron(III)-catecholate complex has sextet spin state and the dioxygen molecule has triplet spin state. Because the DFT method can present reliable energies of the iron(III)-catecholate-dioxygen adduct in the sextet spin state and the separated system in the octet spin state, the reaction energy of the dioxygen adduct formation is evaluated here as the DFT-calculated energy difference between the iron(III)-catecholate-dioxygen adduct in the sextet spin state and the separated system in the octet spin state. This is slightly endothermic relative to the long separated system in the octet spin state.

However, the DFT-calculated PEC of the sextet spin state is much more unstable than the octet and the quartet spin states, in particular, when the Fe-O_2 distance is long (Figure 3(A)). This is because the single reference wave function cannot represent well the sextet spin state when the dioxygen molecule is distant from the iron center; note that the sextet spin state is represented by the combination of the wavefunction of the sextet spin state of the iron(III)-catecholate moiety and the $(\alpha\beta + \beta\alpha)/\sqrt{2}$ wave function of triplet dioxygen molecule (see Scheme 7(A)) at the long Fe-O_2 distance. In summary, the reaction energy is evaluated well with the DFT method, but the multi-reference wavefunction must be employed to investigate the electronic structure change of the sextet spin state, in particular, at the long Fe-O_2 distance.

Considering the above results of the sextet spin state, we tried to evaluate the PEC of

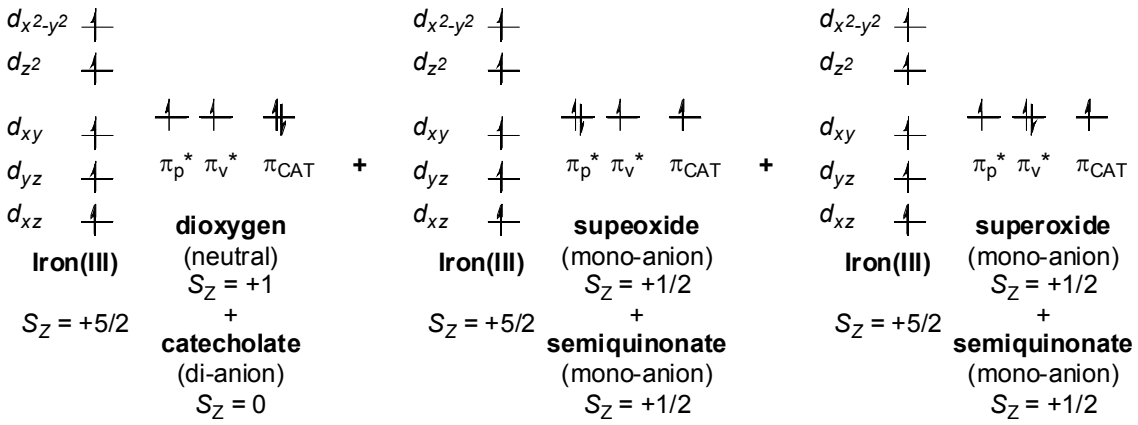


(A) DFT **(B) CASPT2**

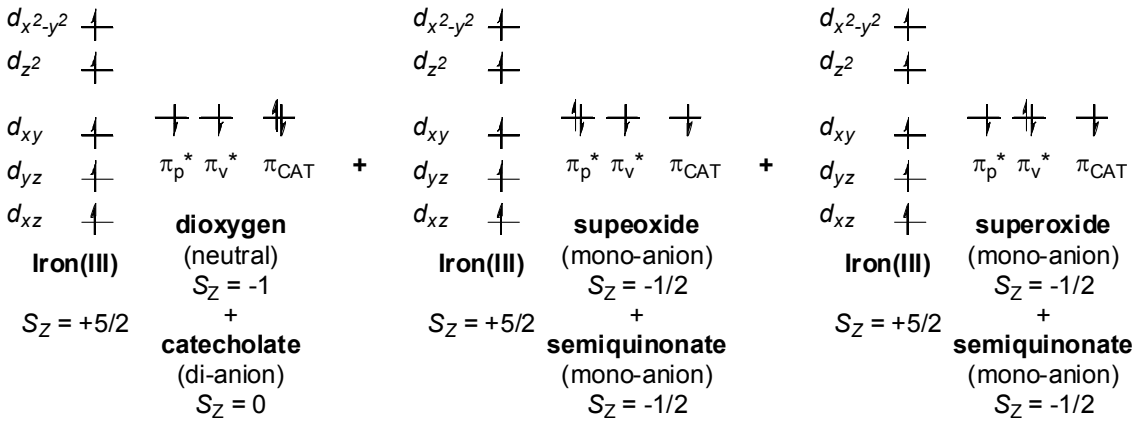
Figure 3. (A) Potential energy curves (PECs) in the O_2 binding process evaluated by the B3LYP/BAS-II method with **SM**, and (B) PECs evaluated by the 3SA-CASSCF(9,8)/CASPT2 method for the octet and quartet spin states and by the 7SA-CASSCF(9,8)/CASPT2 method for the sextet spin state with **SM**. Dashed line, solid line, and dotted line represent octet, sextet, and quartet spin states, respectively. Note that data of R_{Fe-O} = 2.06 Å are evaluated by CASPT2/SS-CASSCF(9,8) method with 0.20 E_h shift

SM with the SS-CASSCF(9,8) method, where the geometry of **SM** was taken to be the same as that of the corresponding moiety of **LM**. However, this method presents unreasonable PEC for all spin states considered here; for instance, the PEC starting from **2a** cannot be connected smoothly with that starting from **1+O_2**, suggesting the electronic state is considerably different between **1+O_2** and **2a**; see Figure A1 for this PEC. This means the SA-CASSCF method should be employed for this O_2 binding process. After preliminary SA-CASSCF calculations with several kinds of averaged states, we evaluated the PEC of **SM**

Scheme 6. Three electronic states considered in the 3SA-CASSCF(9,8) calculation of the octet and quartet spin states



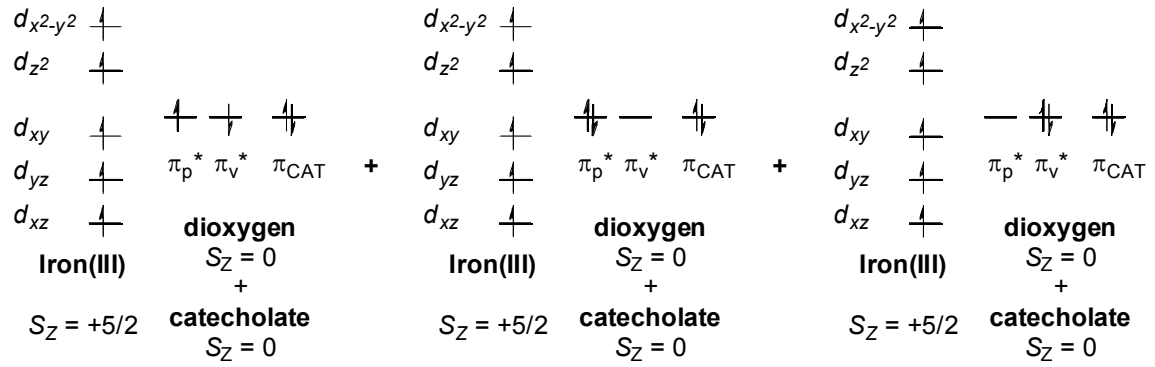
(A) octet spin state (S_Z = +7/2)



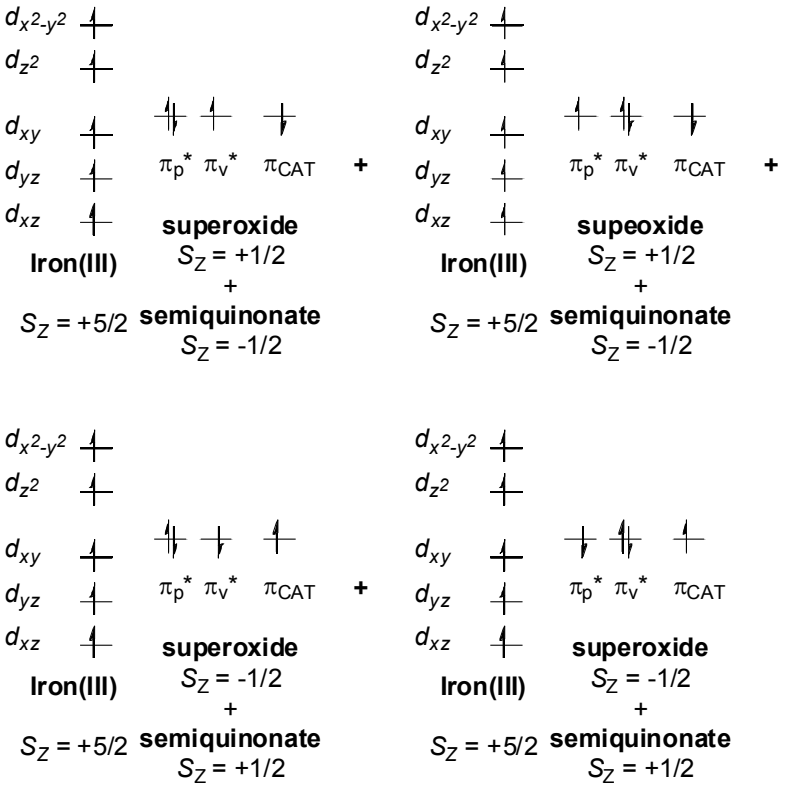
(B) quartet spin state (S_Z = +3/2)

with the 3SA-CASSCF(9,8) for the octet and quartet spin states and the 7SA-CASSCF(9,8) for the sextet spin state; see Figures A1 and A2. In the octet spin state, the 3SA-CASSCF(9,8) calculation contains one electron configuration which corresponds to iron(III) dioxygen catecholate state and two electron configurations which correspond to iron(III) superoxo semiquinonate state, as shown in Scheme 6(A). We carried out the 4SA-CASSCF(9,8) calculation which contains one more electron configuration corresponding

Scheme 7. Seven electronic states considered in the 7SA-CASSCF(9,8) calculation of the sextet spin state



(A) Three electron configurations including singlet spin state of dioxygen moiety



(B) Four kinds of charge transfer from catecholate to dioxygen moiety

to iron(II) dioxygen semiquinonate radical state in addition to the above three electron configurations. However, this iron(II) state little influences the PEC calculated by the 3SA-CASSCF(9,8) because the iron(II) state lies at considerably higher energy than these three iron(III) states; see Figure A1 (C) and (D) for PECs by 3SA-CASSCF(9,8) and 4SA-CASSCF(9,8) calculations. This result indicates that the 3SA-CASSCF(9,8) calculation presents reliable PEC for the octet spin state. In the quartet spin state, the 3SA-CASSCF(9,8) calculation contains three electron configurations like the octet spin state, in which two d electrons of the iron center induce low-spin coupling with either two π^* electrons of neutral dioxygen moiety or one π^* electron of superoxo and one π electron of semiquinonate, as shown in Scheme 6(B). In the sextet spin state, on the other hand, the 3SA-CASSCF(9,8) calculation does not include all the necessary states equally, as follows: When the dioxygen moiety has the singlet spin state locally, three different electron configurations must be considered, as shown in Scheme 7. In addition to these three electron configurations, four electron configurations of charge transfer from the catecholate moiety to the dioxygen moiety must be involved in CASSCF calculation (Scheme 7). Thus, the 7SA-CASSCF(9,8) calculation was necessarily carried out to evaluate the PEC of the sextet spin state. In this 7SA-CASSCF(9,8) calculation, three electron configurations correspond to the iron(III) dioxygen catecholate states and four electron configurations correspond to the iron(III) superoxo semiquinonate radical states.

Thus, the PECs of the octet, sextet, and quartet spin states are calculated by the SA-CASSCF(9,8)/CASPT2 method. As shown in Figure 3(B),[45] the energies of the octet and quartet spin states are very close to each other in all distances of $R_{\text{Fe-O}}$. The sextet spin state is more stable in energy than the octet and quartet spin states, indicating that the sextet is the ground state. This is consistent with the previous theoretical results.[11] Potential energy

monotonously increases in all the octet, sextet, and quartet spin states, as the Fe-O distance decreases from 3.0 Å to 2.4 Å. Then, the energies moderately decrease in these three spin states as the Fe-O distance further decreases from 2.4 Å to **2a** (R_{Fe-O} = 2.06 Å).[46] These PECs are more endothermic and need larger activation barrier than expected from the fact that the enzymatic reaction occurs at room temperature. In other words, the CASPT2 method overestimates the endothermicity and the activation barrier of the O_2 binding process. On the other hand, it should be noted that the CASPT2-calculated energy of the sextet spin state is close to that of the octet spin state at the Fe-O distance of 3.0 Å, as expected. These results, as well as the smooth PEC of the sextet spin state, suggest that the CASPT2 method is useful to investigate how the electronic structure and wavefunction of the sextet spin state change by the approach of dioxygen molecule to the Fe complex.

5.3.4 Electronic Structure Changes in O_2 Binding Process.

Mulliken charges and spin populations of the iron center, catecholate, and dioxygen moieties were evaluated by the 3SA-CASSCF(9,8) method for the octet and quartet spin states and by the 7SA-CASSCF(9,8) method for the sextet spin state, as shown in Figure 4(A) and 4(B), respectively. It should be noted that the Mulliken charge and spin population of the iron center are always 1.1e and 4.8e, respectively, in the octet spin state, 1.1e and 4.2e, respectively, in the sextet spin state, and 1.1e and 4.0e, respectively, in the quartet spin state, during the O_2 binding process. These results clearly indicate that the iron center keeps the high spin d^5 character in the O_2 binding process. This feature is not consistent with the previous proposal in both "substrate activation" and "oxygen activation" mechanisms that the iron(III) center converts to the iron(II) center in the O_2 binding process; in the substrate activation mechanism proposed previously, the high spin iron(III) center converts to high spin

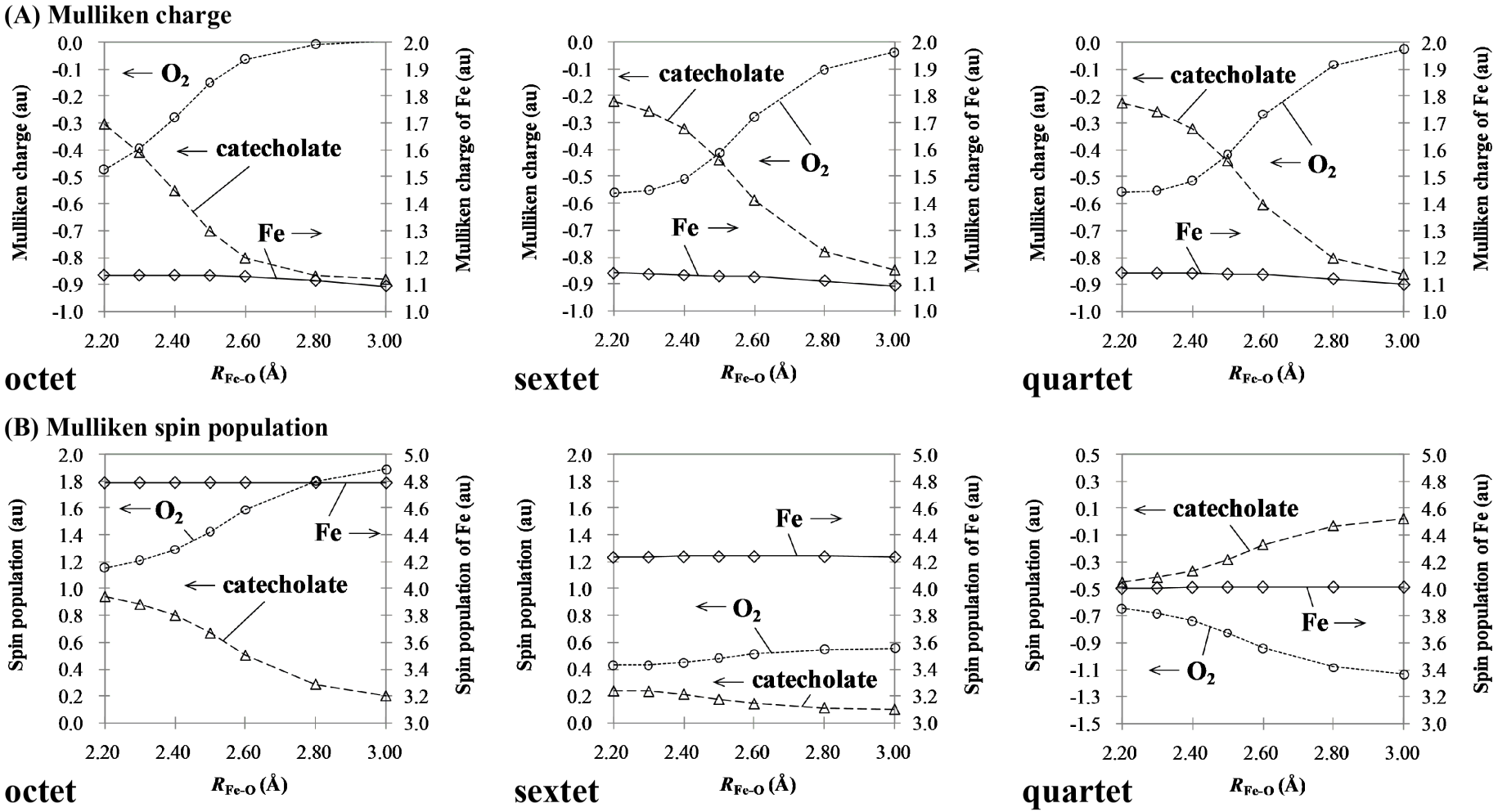


Figure 4. (A) Mulliken charges and (B) Mulliken spin populations with CASSCF method in the O_2 binding process. The 3SA-CASSCF(9,8) method was employed for the octet and quartet spin states and the 7SA-CASSCF(9,8) method was employed for the sextet spin state. Solid line, dashed line, and dotted line represent the Fe center, the catecholate, and the dioxygen moieties, respectively.

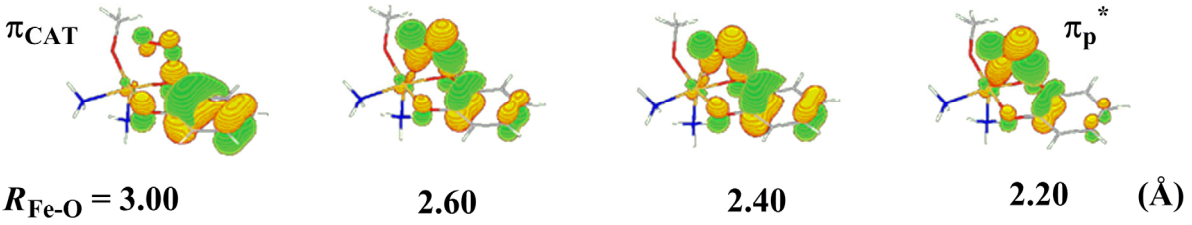


Figure 5. Natural orbital, which possesses occupation number of 2.0, evaluated by the 3SA-CASSCF(9,8) method for octet spin state

iron(II) center by the charge transfer from the catecholate moiety to the iron(III) center to yield reactive semiquinonate species which possesses radical center on the catechol ring. In the oxygen activation mechanism proposed previously, the charge transfer occurs from the catecholate moiety to the iron(III) center, too, so as to easily induce the charge transfer to the dioxygen moiety from the iron center.

Though the Mulliken charge and spin population of the iron center little change, the spin population of the catecholate moiety moderately increases and that of the dioxygen moiety moderately decreases in the sextet spin state as the O_2 binding proceeds (Figure 4(B)). At the same time, the Mulliken charge of the catecholate moiety gradually decreases and that of the dioxygen moiety gradually increases (Figure 4(A)). In the octet and quartet spin states, the Mulliken charges change similar to those in the sextet spin state. Though the spin population of the catecholate moiety considerably increases and that of the dioxygen moiety considerably decreases in the octet spin state, that of the catecholate moiety considerably decreases and that of the dioxygen moiety considerably increases in the quartet spin state. It is noted that though the spin population of the dioxygen moiety moderately increases and that of the catecholate moiety moderately decreases in the sextet spin state unlike those in the octet and quartet spin states (Figure 4(B)), Mulliken charges considerably change in the sextet

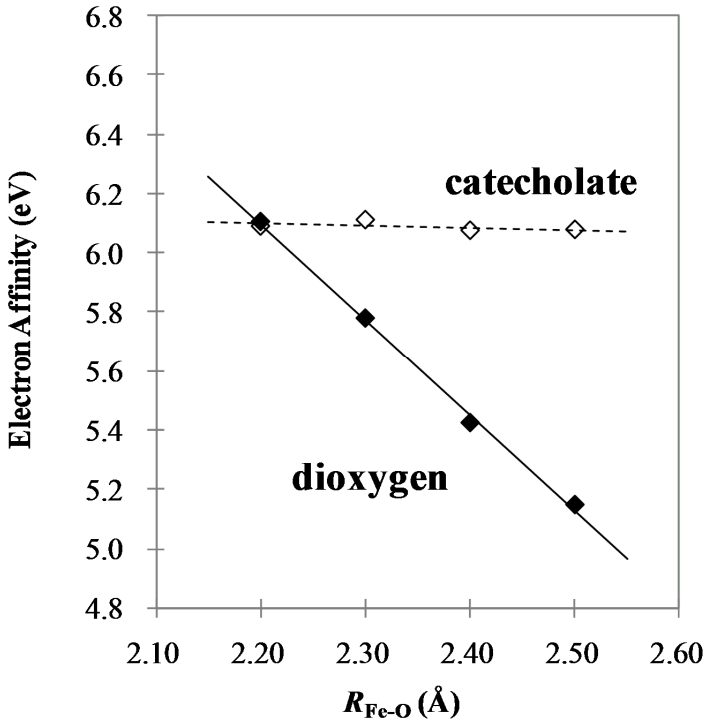


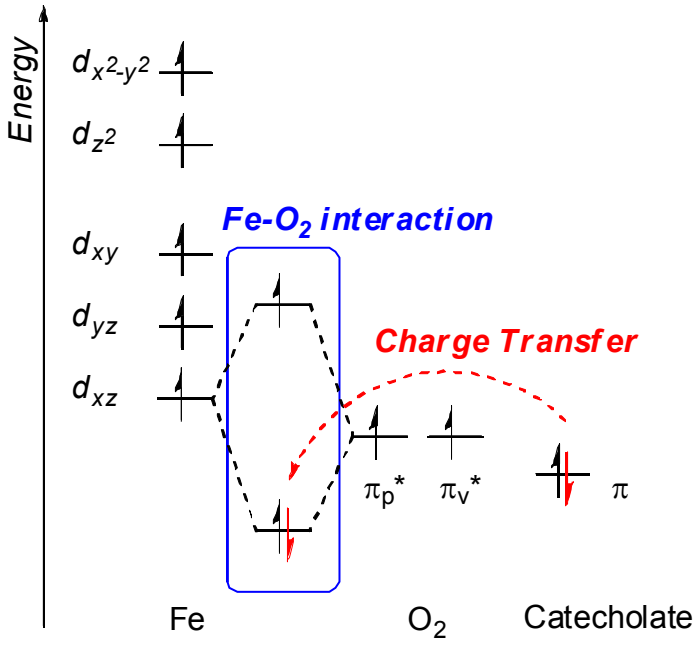
Figure 6. Vertical electron affinity of the dioxygen moiety (black) and the catecholate moiety (red). In the catecholate moiety, the electron affinity of $\mathbf{1}_{SM}$ was evaluated by the CASPT2/SS-CASSCF(6,6) and the CASPT2/SS-CASSCF(7,6) methods for the septet and sextet spin states, which correspond to the iron(III)-semiquinonate and the iron(III)-catecholate states, respectively. In the dioxygen moiety, the electron affinity of $\mathbf{2}_{SM}$ were evaluated by the CASPT2/SS-CASSCF(8,8) and the CASPT2/SS-CASSCF(9,8) methods for the nonet and octet spin states, which correspond to the iron(III)-semiquinonate-neutral dioxygen and iron(III)-semiquinonate-superoxo states, respectively.

spin state similar to those of the octet and quartet spin states (Figure 4(A)). These results are consistent with the electron configurations shown in Scheme 7; in the sextet spin state, four kinds of electron configuration (Scheme 7(B)) participate in the charge transfer from the catecholate moiety to the dioxygen moiety, in which spin populations of the catecholate and dioxygen moieties increase in two electron configurations but decrease in the other two electron configurations. Also, the little change in Mulliken charge and spin population of the iron center indicate that the charge transfer directly occurs from the catecholate dianion taking closed-shell singlet state to the dioxygen moiety taking triplet spin state (Scheme 7), to afford finally superoxide and semiquinonate monoanion which both have doublet spin state locally.

One doubly occupied natural orbital is responsible for the changes in spin and electron populations, as follows; the doubly occupied natural orbital is localized in the π orbital of the catecholate moiety at $R_{Fe\text{-}O}$ = 3.0 Å, as shown in Figure 5. The π orbital of the catecholate moiety starts to interact with the π_p* orbital of the dioxygen moiety at $R_{Fe\text{-}O}$ = 2.4 ~ 2.6 Å, and finally, the doubly occupied orbital becomes localized in the π_p* orbital of the dioxygen moiety at $R_{Fe\text{-}O}$ = 2.2 Å. These features suggest that the π orbital of the catecholate moiety and the π_p* orbital of the dioxygen moiety participate in the direct charge transfer, which is consistent with the result that the Mulliken charge and spin population of the iron(III) center little change in the reaction. Interestingly, the d orbital of the iron center little participates in this orbital. This is against our expectation that the iron center plays important role in the dioxygen activation. It is of considerable importance to clarify how the iron(III) center participates in the dioxygen activation and why the iron(III) center is necessary in the O_2 binding process.

To investigate the role of the iron(III) center, we evaluate vertical electron affinities

Scheme 8. Fe-O_2 interaction induces the charge transfer from the catecholate moiety to the dioxygen moiety



($E_{EA,v}$) of the catecholate and dioxygen moieties with the CASPT2 method. In the evaluation of $E_{EA,v}$ of the catecholate moiety, a model complex which consisted of iron center, catecholate moiety, two ammonia ligands, and methoxide was employed. This is similar to $\mathbf{1}_{SM}$. The SS-CASSCF(6,6)/CASPT2 and SS-CASSCF(7,6)/CASPT2 methods were applied to the septet and sextet spin states, respectively. In the evaluation of $E_{EA,v}$ of the dioxygen moiety, a model complex which consisted of iron center, catecholate moiety, dioxygen moiety, two ammonia ligands, and methoxide was employed. This is similar to $\mathbf{2a}_{SM}$. The SS-CASSCF(8,8)/CASPT2 and SS-CASSCF(9,8)/CASPT2 methods were applied to the nonet and octet spin states, respectively. The $E_{EA,v}$ of the catecholate moiety is always 6.1 eV, which little changes during the O_2 binding, as shown in Figure 6. On the other hand, the $E_{EA,v}$ of the dioxygen moiety considerably increases as the dioxygen molecule approaches the iron center. It is noted that the $E_{EA,v}$ of the dioxygen moiety becomes larger than that of the catecholate moiety, when the Fe-O distance becomes shorter than 2.2 Å. From these results,

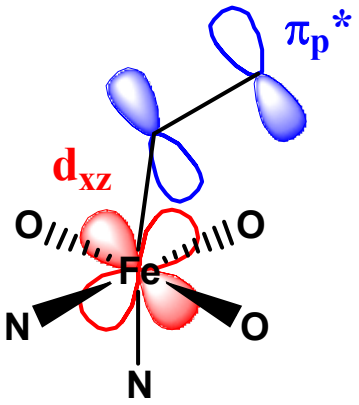
it should be concluded that the role of iron center is "adjusting" the orbital energy of the dioxygen moiety to induce charge transfer from the catecholate moiety to the dioxygen moiety, which is schematically showin in Scheme 8.

5.3.5 η^2-Side-on Coordination Superoxo Complex 2b.

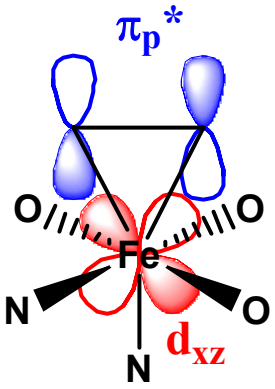
Unexpectedly, the side-on type iron(III)-dioxygen complex $\mathbf{2b_{LM}}$ was optimized by the DFT method in the octet, sextet, and quartet spin states (Figure 1(C)). This type of iron(III) η^2-side-on dioxygen complex has not been reported yet both theoretically and experimentally. The geometry is little different in these three spin states. The O3-O4 bond length is 1.288-1.304 Å, which is in the range of superoxide like that of $\mathbf{2a_{LM}}$. The Fe-O3 distance is 2.170-2.196 Å and the Fe-O4 distance is 2.143-2.196 Å, indicating that the η^2-side-on coordination mode is stable without any support such as hydrogen bond. The sextet and quartet spin states are calculated with the DFT method to be less stable than the octet spin state by 5.5 and 5.6 kcal/mol, respectively.

The SS-CASSCF(9,8)-computed Mulliken spin populations are 4.79e, 1.04e, and 1.06e for the iron center, catecholate, and dioxygen moieties, respectively, in the octet spin state, 4.55e, -0.67e, and 1.01e, respectively, in the sextet spin state, and 4.02e, -0.56e, and -0.55e, respectively, in the quartet state. From these results, it is concluded that **2b** should be characterized to be an iron(III)-superoxo complex, which is consistent with the geometrical features discussed above. The doubly occupied natural orbital involves bonding interaction between the d_{xz} orbital of Fe and the π_p^* orbital of the dioxygen moiety, as shown in Figure 2(B). Though the d_{xz} orbital of Fe interacts with the π_p^* orbital in both of $\mathbf{2a_{SM}}$ and $\mathbf{2b_{SM}}$, bonding interaction is different between **2a** and **2b**. In **2a**, one lobe of the d_{xz} orbital overlaps with one lobe of the π_p^* orbital in a bonding way, as shown in Scheme 9(A), which

Scheme 9. Bonding interaction between the d_{xz} orbital of the Fe center and the π_p* orbital of the dioxygen moiety



(A) End-on Complex 2a



(B) Side-on complex 2b

is the typical bonding interaction of the η^1-end-on type coordination of dioxygen molecule. In **2b**, on the other hand, two lobes of the d_{xz} orbital overlap well with two lobes of the π_p* orbital in a bonding way (see Scheme 9(B)), which is the typical bonding interaction of the usual η^2-side-on coordination of dioxygen molecule.

In the SS-CASSCF(9,8)/CASPT2 calculation, $\mathbf{2b_{SM}}$ is more stable than $\mathbf{2a_{SM}}$ by 9.1, 12.6, and 12.7 kcal/mol and more unstable than $\mathbf{1+O_2}$ by 13.0, 6.2, and 8.9 kcal/mol, in the octet, sextet, and quartet spin states, respectively. Also, the sextet spin state of $\mathbf{2b_{SM}}$ is calculated to be more stable than the octet and quartet spin states by 6.8 and 2.7 kcal/mol, respectively. Thus, the ground state is the sextet spin state in both $\mathbf{2a_{SM}}$ and $\mathbf{2b_{SM}}$. In the DFT calculation with the same basis sets and the same geometry as those of the SS-CASSCF(9,8)/CASPT2 calculation, $\mathbf{2b_{SM}}$ is more stable than $\mathbf{2a_{SM}}$ by 4.1, 5.7, and 0.7 kcal/mol, in the octet, sextet, and quartet spin states, respectively. Unlike the SS-CASSCF(9,8)/CASPT2 calculation, $\mathbf{2b_{SM}}$ is more stable than $\mathbf{1+O_2}$ by 2.9 and 2.6 kcal/mol in the octet and sextet spin states, respectively, but more unstable than $\mathbf{1+O_2}$ by 2.3 kcal/mol in the quartet spin state. Thus, the ground state is either sextet or octet spin state in

2b$_{SM}$, according to the DFT calculations.

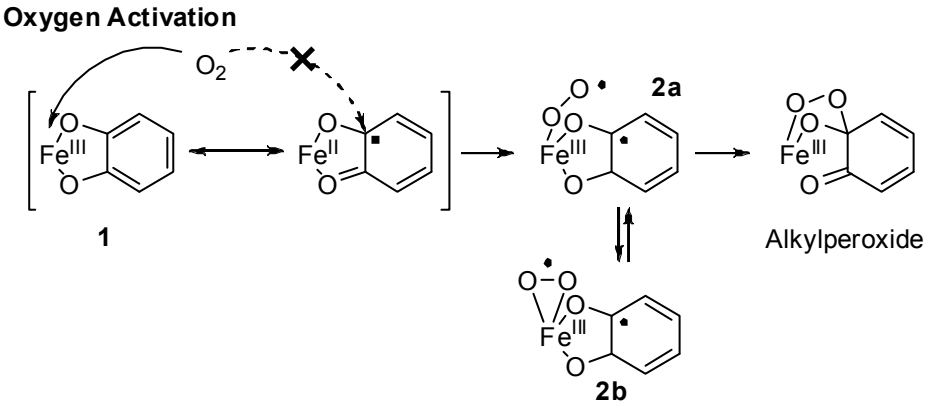
Both SS-CASSCF(9,8)/CASPT2- and DFT-computed results indicate that **2b** is a stable intermediate. It is likely that **2b** exists as a stable isomer in the early stage of the catalytic cycle of iron(III) catechol dioxygenase.

5.4 Conclusions

We carried out here the CASSCF/CASPT2 study of the dioxygen binding process in iron(III) catechol dioxygenase. Our results clearly support the "oxygen activation" mechanism and rule out the "substrate activation" mechanism, because semiquinonate radical character is little involved in the iron(III) catecholate complex.

We wish to present the clear feature of the dioxygen activation, as summarized in Scheme 10. First, iron(III) catecholate complex **1** is formed. In the ground state, the iron center is characterized to be high-spin d^5 species, and semiquinonate radical character is little observed, which is not enough to induce the direct reaction between the dioxygen molecule and the catechol ring. In the second step, the dioxygen molecule directly attacks the iron center to form the η^1-end-on type Fe-O_2 complex **2a** which is understood to be iron(III) superoxo complex. O_2 binding process was investigated by the CASSCF/CASPT2 method. When dioxygen molecule is distant from the iron(III) center, the octet, sextet, and quartet spin states are nearly degenerate in the iron(III) dioxygen system (**1+O$_2$**). The charge transfer from the catecholate moiety to the dioxygen moiety directly occurs without formation of iron(II) character, when the Fe-O_2 distance becomes 2.4 Å or less. This is the key process of the dioxygen activation. Around the Fe-O_2 distance of 2.2 Å, the sextet spin state becomes more stable than the octet and quartet spin states. Finally, the iron(III)-η^1-end-on superoxo species is formed, spin state of which is sextet. In this process, the Mulliken charge and spin

Scheme 10. Proposed mechanism of the dioxygen activation in iron(III) catechol dioxygenase



population of the iron center little change. However, these results do not mean that the iron center does not play important role in the dioxygen activation process. The vertical electron affinity of the dioxygen moiety was calculated to be larger than that of the catecholate moiety where the Fe-O distance becomes short. This indicates that the iron center induces the charge transfer from the catecholate moiety to the dioxygen molecule by adjusting the energy level of the π_p* orbital of O_2.

Besides **2a**, we found the iron(III)-η^2-side-on superoxo complex **2b** as a stable intermediate, which has been neither observed nor predicted in both experimentally and theoretically. These results suggest that **2b** exists in a stable isomer of earlier stage of the catalytic cycle by iron(III) catechol dioxygenase.

5.5 Appendix

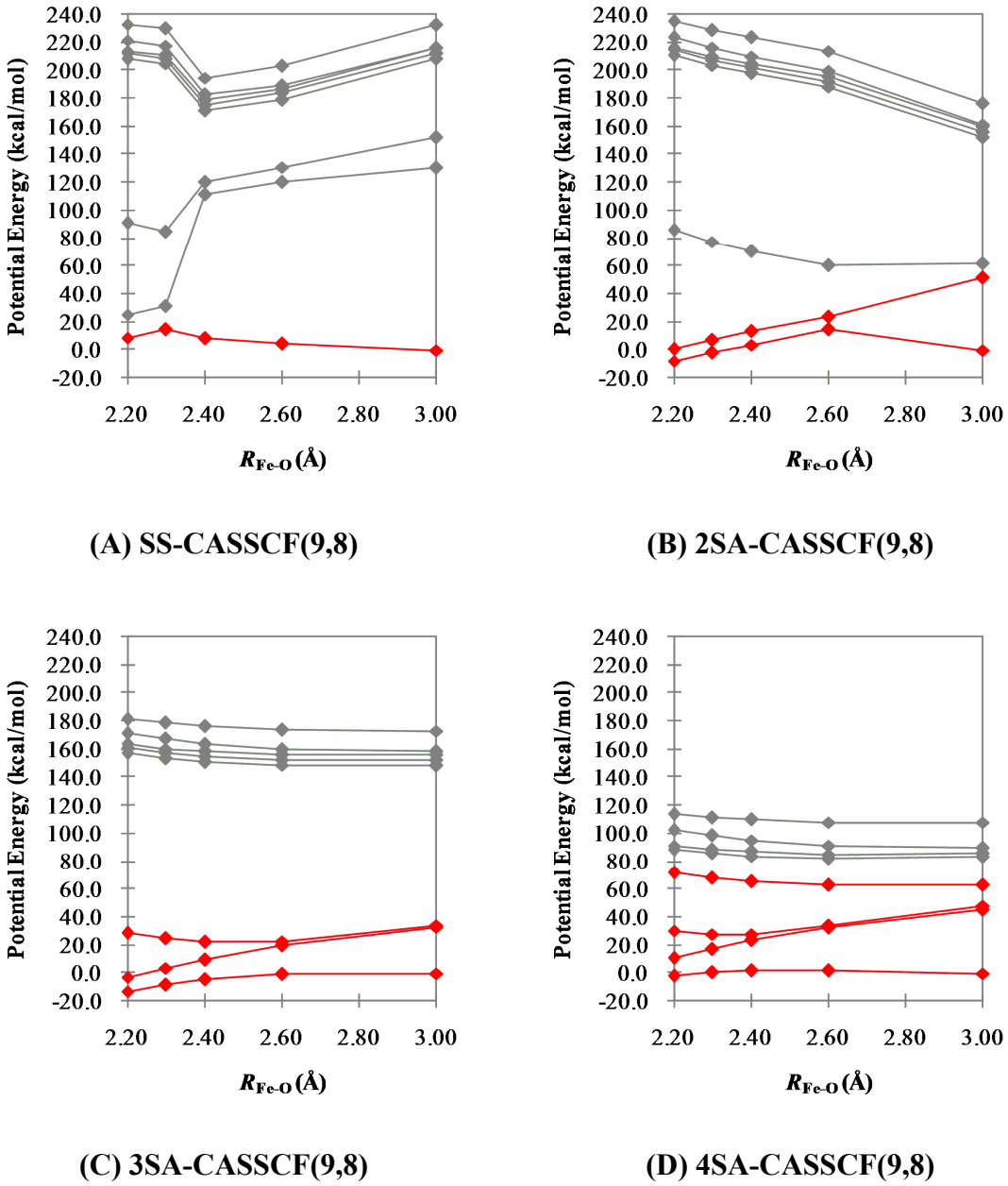
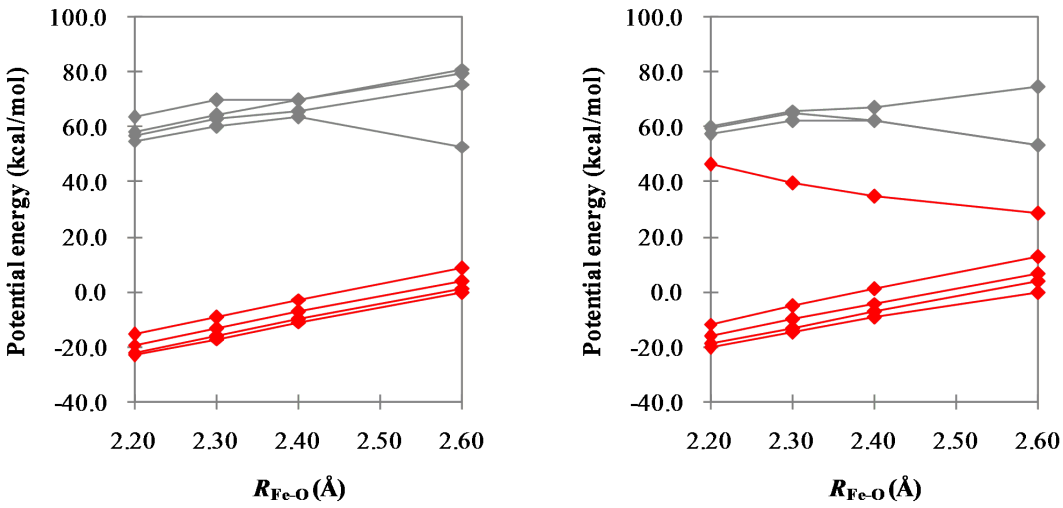
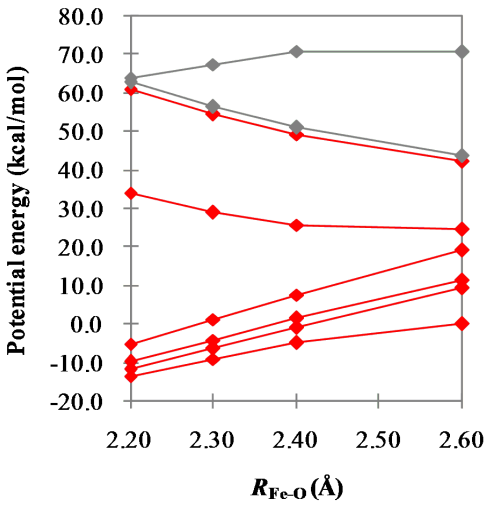


Figure A1. Potential energy curves with SS-CASSCF(9,8) and several SA-CASSCF(9,8) methods in the octet spin state. Red lines represent averaged-states and gray lines represent other states

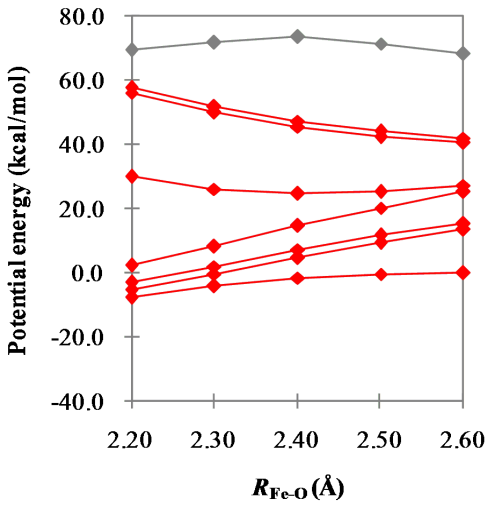


(A) 4SA-CASSCF(9,8)

(B) 5SA-CASSCF(9,8)



(C) 6SA-CASSCF(9,8)



(D) 7SA-CASSCF(9,8)

Figure A2. Potential energy curves with several SA-CASSCF(9,8) methods in the sextet spin state. Red lines represent averaged-states and gray lines represent other states

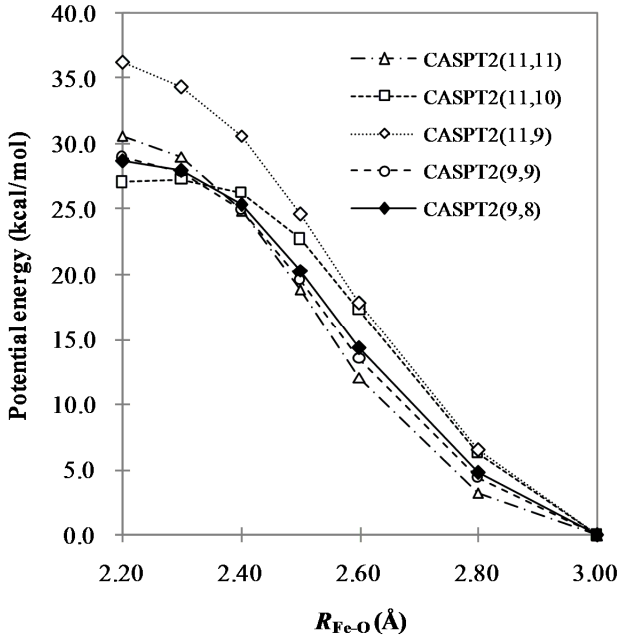


Figure A3. PECs in the octet spin state calculated with several CASSCF/CASPT2 calculations. CASPT2(X,Y) means 3SA-CASSCF(X,Y)/LS-CASPT2 method with the level shift of 0.30 E_h

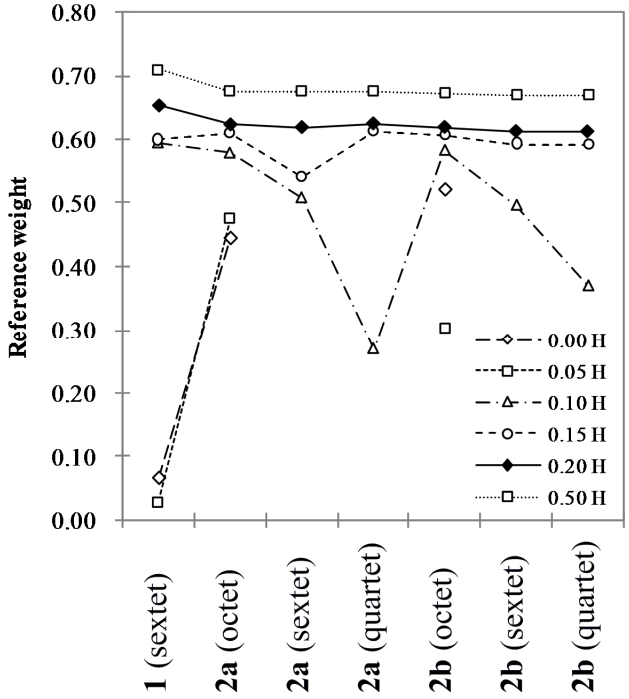


Figure A4. Reference weights calculated by the LS-CASPT2 method with several level shifts. In no shift and very small shift value of 0.05 E_h, some of data could not be presented because the CASPT2 iteration was not converged, indicating that there are strong intruder states and the reference weights are too small in some calculations. Note that the reference weights are not constant where the shift value is less than 0.20 E_h, because the intruder states are not completely reduced with such shift value.

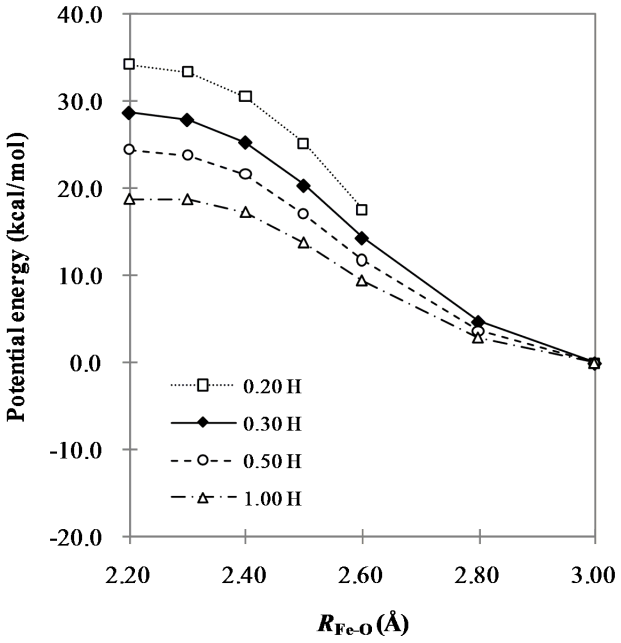


Figure A5. The PECs of the octet spin state calculated with the 3SA-CASSCF(9,8)/LS-CASPT2 method using several level shifts. Note that the energy with shift value of 0.20 E_h at R_{Fe-O} = 2.80 Å is omitted here, because the perturbation theory is broken at this point.

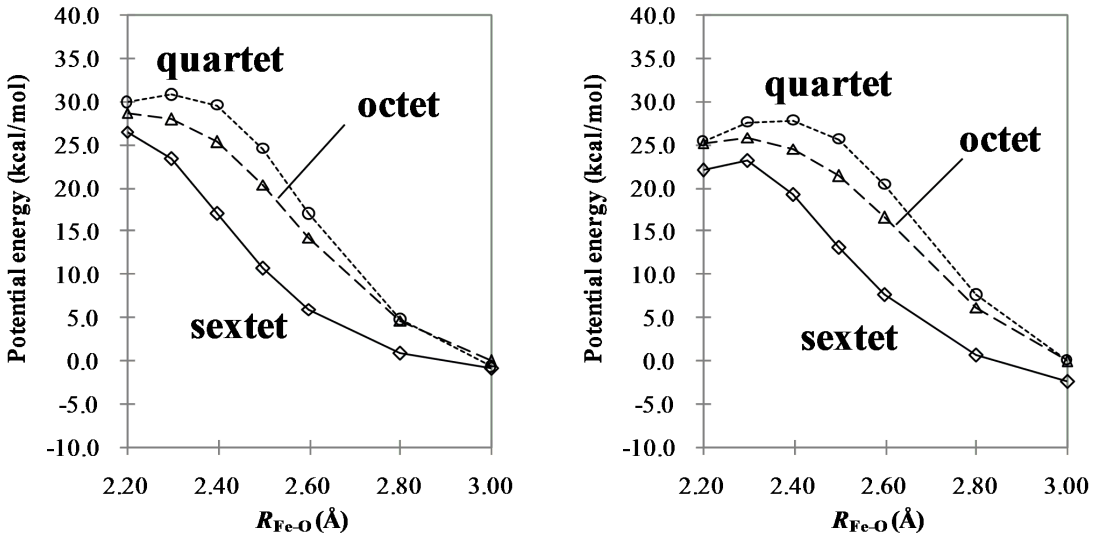


Figure A6. PECs of the octet, sextet, and quartet spin states taking the octet and sextet geometries, where the DFT method was employed for geometry optimization. The 3SA-CASSCF(9,8)/LS-CASPT2 method was applied to the octet and quartet spin states with the shift of 0.30 E_h, and the 7SA-CASSCF(9,8)/LS-CASPT2 method was applied to the sextet spin state with the shift of 0.30 E_h.

References

(1) Gibson, D. T. In *Microbial Degradation of Organic Compounds*; Marcel Dekker: New York, **1984**.

(2) Lipscomb, J. D.; Orville, A. M. In *Metal Ions in Biological Systems*; Sigel, H., Sigel, A., Eds.; Marcel Dekker: New York, **1992**; Vol. 28, pp 243-298.

(3) Funabiki, T. In *Oxygenases and Model Systems*; Funabiki, T., Ed.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1997; pp19-104.

(4) Costas, M.; Mehn, M. P.; Jensen, M. P.; Que, L., Jr. *Chem. Rev.* **2004**, *104*, 939-986.

(5) Bugg, T. D. H. *Tetrahedron* **2003**, *59*, 7075-7101.

(6) Hayaishi, O.; Katagiri, M.; Rothberg, S. *J. Am. Chem. Soc.* **1955**, *77*, 5450-5451.

(7) Kojima, Y.; Itada, N.; Hayaishi, O. *J. Biol. Chem.* **1961**, *236*, 2223-2228.

(8) Ohlendorf, D. H.; Lipscomb, J. D.; Webber, P. C. *Nature* **1988**, *336*, 403-405.

(9) Ohlendorf, D. H.; Orville, A. M.; Lipscomb, J. D. *J. Mol. Biol.* **1994**, *244*, 429-440.

(10) Orville, A. M.; Lipscomb, J. D.; Ohlendorf, D. H. *Biochemistry* **1997**, *36*, 10052-10066.

(11) Que, L., Jr.; Lipscomb, J. D.; Zimmermann, R.; Muenck, E.; Orme-Johnson, N. R.; Orme-Johnson, W. H. *Biochim. Biophys. Acta* **1976**, *452*, 320-334.

(12) Que, L., Jr.; Heistand, R. H., II. *J. Am. Chem. Soc.* **1979**, *101*, 2219-2221.

(13) Walsh, T. A.; Ballou, D. P.; Mayer, R.; Que, L., Jr. *J. Biol. Chem.* **1983**, *258*, 14422-14427.

(14) Whittaker, J. W.; Lipscomb, J. D.; Kent, T. A.; Muenck, E. *J. Biol. Chem.* **1984**, *259*, 4466-4475.

(15) Funabiki, T.; Mizoguchi, A.; Sugimoto, T.; Tada, S.; Tsuji, M.; Sakamoto, H.; Yoshida, S. *J. Am. Chem. Soc.* **1986**, *108*, 2921-2932.

(16) Que, L., Jr.; Lipscomb, J. D.; Muenck, E.; Wood, J. M. *Biochim. Biophys. Acta* **1977**, *485*,

60-74.

(17) Que, L., Jr.; Kolanczyk, R. C.; White, L. S. *J. Am. Chem. Soc.* **1987**, *109*, 5373-5380.

(18) Cox, D. D.; Que, L., Jr. *J. Am. Chem. Soc.* **1988**, *110*, 8085-8092.

(19) Funabiki, T.; Fukui, A.; Hitomi, Y.; Higuchi, M.; Yamamoto, T.; Tanaka, T.; Tani, F.; Naruta, Y. *J. Inorg. Biochem.* **2002**, *91*, 151-158.

(20) Hitomi, Y.; Yoshida, M.; Higuchi, M.; Minami, H.; Tanaka, T.; Funabiki, T. *J. Inorg. Biochem.* **2005**, *99*,755-763.

(21) Higuchi, M.; Hitomi, Y.; Minami, H.; Tanaka, T.; Funabiki, T. *Inorg. Chem.* **2005**, *44*, 8810-8821.

(22) Bull, C.; Ballou, D. P.; Otsuka, S. *J. Biol. Chem.* **1981**, *256*, 2681-2686.

(23) Hitomi, Y.; Tase, Y.; Higuchi, M.; Tanaka, T.; Funabiki, T. *Chem. Lett.* **2004**, *33*, 316-317.

(24) Funabiki, T.; Inoue, T.; Kojima, H.; Konishi, T.; Tanaka, T.; Yoshida, S. *J. Mol. Catal.* **1990**, *59*, 367-371.

(25) Funabiki, T.; Yamazaki, T. *J. Mol. Catal. A.* **1999**, *150*, 37-47.

(26) Borowski, T.; Siegbahn, P. E. M. *J. Am. Chem. Soc.* **2006**, *128*, 12941-12953.

(27) Pau, M. Y. M.; Davis, M. I.; Orville, A. M.; Lipscomb, J. D.; Solomon, E. I. *J. Am. Chem. Soc.* **2007**, *129*, 1944-1958.

(28) Roos, B. O.; Taylor, P. R. *Chem. Phys.* **1980**, *48*, 157-173.

(29) Roos, B. O. In *Ab Initio Methods in Quantum Chemistry II*; Lawley, K. P., Ed.; Wiley: New York, **1987**, p.399.

(30) Andersson, K.; Malmqvist, P.-Å.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. *J. Phys. Chem.* **1990**, *94*, 5483-5488.

(31) Andersson, K.; Malmqvist, P.-Å.; Roos, B. O. *J. Chem. Phys.* **1992**, *96*, 1218-1226.

(32) Arginine residue (Arg457) exists close to the active site. It plays important role in the

conformational change of an alkylperoxo complex which occurs after the dioxygen activation process.[26] However, recent theoretical study reported that the arginine residue little participates in the dioxygen activation process.[26] Thus, we did not consider Arg457 in our calculation models.

(a) To estimate protein environmental effects, the energies and electronic structures were also evaluated with PCM (polarized continuum model) method[32b] at the DFT (B3LYP) level, employing dielectric constant of 4.0 and probe radius of 1.4 Å for **LM**. These results showed that the relative energies and electronic structures are little changed by the protein environment in the dioxygen activation process. (b) Cancès, M. T.; Mannucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032-3041.; Cossi, M.; Barone, V.; Mennucci, B.; Tomasi, J. *Chem. Phys. Lett.* **1998**, *286*, 253-260.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *106*, 5151-5158.

(33) Lee, C.; Yang, W.; Parr, R.G. *Phis. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785-789.

(34) Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. *Chem. Phys. Lett.* **1989**, *157*, 200-206.

(35) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648-5652.

(36) Dolg, M.; Wedig, U.; Stoll, H.; Preuss, H. *J. Chem. Phys.* **1987**, *86*, 866-872.

(37) Roos, B. O.; Andersson, K. *Chem. Phys. Lett.* **1995**, *245*, 215-223.

(38) Roos, B. O.; Andersson, K.; Fülscher, M. P.; Serrano-Andrés, L.; Pierloot, K.; Merchán, M.; Molina, V. *J. Mol. Struct. Theochem.* **1996**, *388*, 257-276.

(39) Frisch, M. J.; et al. *Gaussian 03*, Revision C.02; Gaussian Inc.; Wallingford, CT, 2004.

(40) Karlström, G.; Lindh, R.; Malmqvist, P.-Å.; Roos, B. O.; Ryde, U.; Veryazov, V.; Widmark, P.-O.; Cossi, M.; Schimmelpfennig, B.; Neogrady, P.; Seijo, L. *Comput. Mater. Sci.* **2003**, *28*, 222-239.

(41) Herzberg, G. In *Spectra of Diatomic Molecules*; Spinks, J. W. D.; 2nd Ed.; Krieger Pub. Co.: **1989**.

(42) Cramer, C. J.; Tolman, W. B.; Theopold, K. H.; Rheingold, A. L. *Proc. Nat. Acad. Sci. U.S.A.* **2003**, *100*, 3635-3640.

(43) The π_v orbital of the dioxygen moiety was involved in the active space instead of the π_p* orbital in the sextet spin state. However, this is not unreasonable, as follows: Because both natural orbitals (NOs) possess occupation number of 2.00, these two NOs little contribute to the correlation energy of SS-CASSCF(9,8) calculation, and therefore, the orbital order accidentally changes between these NOs.

(44) First, PECs of the octet, sextet, and quartet spin states were calculated with the DFT-optimized geometry in each spin state. However, the octet and sextet spin states calculated with the quartet geometry are more stable than those with the octet and sextet geometries. This suggests that true geometries of $\mathbf{2a_{SM}}$ in the octet and sextet spin states are more close to the DFT-optimized geometry for the quartet spin state rather than the DFT-optimized geometries for the octet and sextet states. Thus, we employed the DFT-optimized geometry for the quartet spin state to evaluate PECs of the octet and sextet spin states and to discuss their energies and electronic structures. PECs calculated with the DFT-optimized geometries for the octet and sextet spin states are presented in Figure A6.

(45) We did not present the PECs between $R_{Fe\text{-}O}$ = 2.20 Å and the DFT-optimized distance of **2a** in the sextet spin state ($R_{Fe\text{-}O}$ = 2.06 Å) in Figure 3, because the energy of $\mathbf{2a_{SM}}$ could not be reliably computed by the 7SA-CASSCF(9,8)/CASPT2 method, as follows: In $\mathbf{2a_{SM}}$, higher two states in the seven electron configurations considered by 7SA-CASSCF(9,8) calculation become more unstable and cannot be involved in

7SA-CASSCF(9,8) calculation. Instead of 7SA-CASSCF(9,8) calculation, we performed SS-CASSCF(9,8)/CASPT2 calculations to examine PEC between R_{Fe-O} = 2.20 Å and 2.06 Å, because SS-CASSCF(9,8)/CASPT2 method presents reliable energy between R_{Fe-O} = 2.06 Å and 2.30 Å (see above). The SS-CASSCF(9,8)/CASPT2-calculated PEC slightly decreases upon going from R_{Fe-O} = 2.20 Å to 2.06 Å, suggesting that **2a** is in shallow minimum and the activation barrier is near to the energy difference between R_{Fe-O} = 3.00 Å and 2.20 Å, which is about 20.5 kcal/mol.

Chapter 6

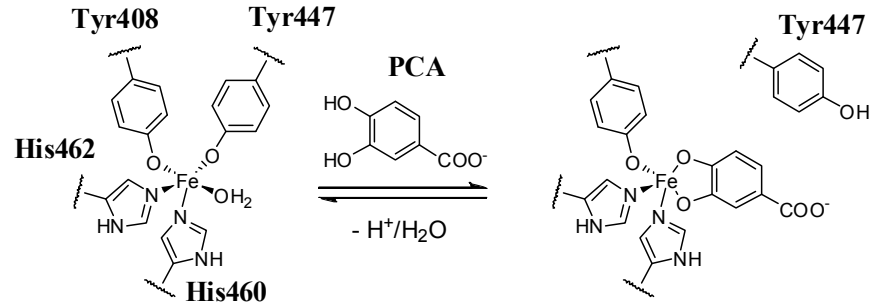
Multi-State CASPT2 Study of Native Iron(III)-dependent Catechol Dioxygenase and Its Functional Models: Comparisons of Electronic Structure and Ligand-to-Metal Charge-Transfer Excitation

6.1 Introductions

Oxygenase enzymes perform various biodegradation and biosynthetic reactions with dioxygen molecule (O_2). It is well known that dioxygen molecule takes a triplet spin state in the ground state, which is not reactive to the most of organic molecules. Thus, activation of dioxygen molecule is essential to the function of oxygenase enzymes. Catechol dioxygenase, found in wide range of soil bacteria, activates dioxygen molecule to catalyze aromatic bond cleavage of catechol (1,2-hydroxybenzene) and its derivatives.[1] This oxygenation is important as the last step of biodegradation of aromatic compounds. Catechol dioxygenase includes either non-heme iron(II) or iron(III) complex in its active site which plays key role in dioxygen activation process. In this regard, a lot of studies including theoretical studies were carried out as summarized in recent reviews of this subject.[2-6] In particular, iron(III)-dependent catechol dioxygenase is of considerable interest because its oxygenation reaction is against our general understanding that a high-spin iron(III) center does not react with dioxygen molecule.

Protocatechuate 3,4-dioxygenase (PCD), the most typical example of iron(III)-dependent catechol dioxygenase, contains a high-spin iron(III) center coordinated with two histidines (His460 and His462), two tyrosines (Tyr408 and Tyr447), and one water

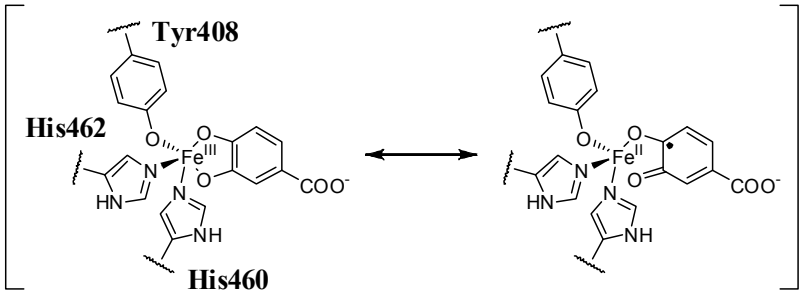
Scheme 1



molecule in its active site.[7] The substrate protocatechuate (3,4-dihydroxybenzoate, PCA) is bound with the active center to form a PCD—PCA complex **1** with concomitant dissociation of Tyr447 and water molecule,[8] as shown in Scheme 1. The active center involves 5-coordinate catecholatoiron(III) complex. Because a high-spin iron(III) center does not react with a dioxygen molecule in general, the iron(III)—catecholate form (represented by iron(III)—CAT, hereafter)[9] is believed to be inactive for dioxygen activation. However, the PCD rapidly catalyzes oxygenation reaction of PCA with a large rate constant of 5×10^5 M^{-1} s^{-1}.[10]

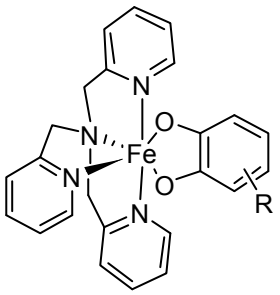
On the other hand, large numbers of bio-mimetic and bio-inspired catecholatoiron(III) complexes have been synthesized for these two decades.[2-6] Interestingly, some of those complexes exhibit the catalytic activity with dioxygen molecule to induce aromatic bond cleavage.[11] They are *so called* functional model complexes. Dioxygen activation process by iron(III)-dependent catechol dioxygenase is often discussed based on those functional model complexes.[2-6] Because 6-coordinate catecholatoiron(III) model complex reacts with dioxygen molecule despite the absence of the coordinating site,[12] it has been proposed that dioxygen molecule directly reacts with the catecholate moiety. This proposal has been accepted as the "substrate activation" mechanism.[12] Because closed-shell singlet catecholate anion is not reactive for dioxygen molecule, it is likely that iron(II)—sqmiquinonate complex

Scheme 2

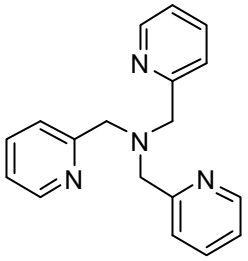


iron(III)—catecholate **iron(II)—semiquinonate**

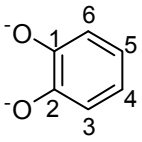
Scheme 3



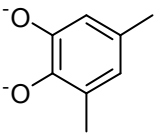
[Fe^{III}(TPA)(R-CAT)]



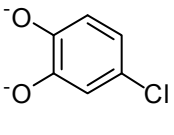
TPA = tris(2-pyridylmethyl)amine



CAT = catecholate



DMC = 3,5-dimethylcatecholate



4CC = 4-chlorocatecholate

(represented by iron(II)—SQ, hereafter)[9] is formed through the ligand-to-metal charge transfer (LMCT) from the catecholate moiety to the iron(III) center, and its radical center on the catechol ring reacts with the dioxygen molecule. Actually, the LMCT absorption is observed in near-infrared region both in the native catechol dioxygenase and in the functional models. For instance, the PCD-PCA complex **1** exhibits the LMCT excitation in the visible light region (1.65 eV).[13] However, the energy gap is still too large to induce thermal

excitation to the LMCT excited state. Thus, it is discussed in experimental chemistry that the iron(II)—SQ state participates in the ground state through *resonance structure* (Scheme 2).[14]

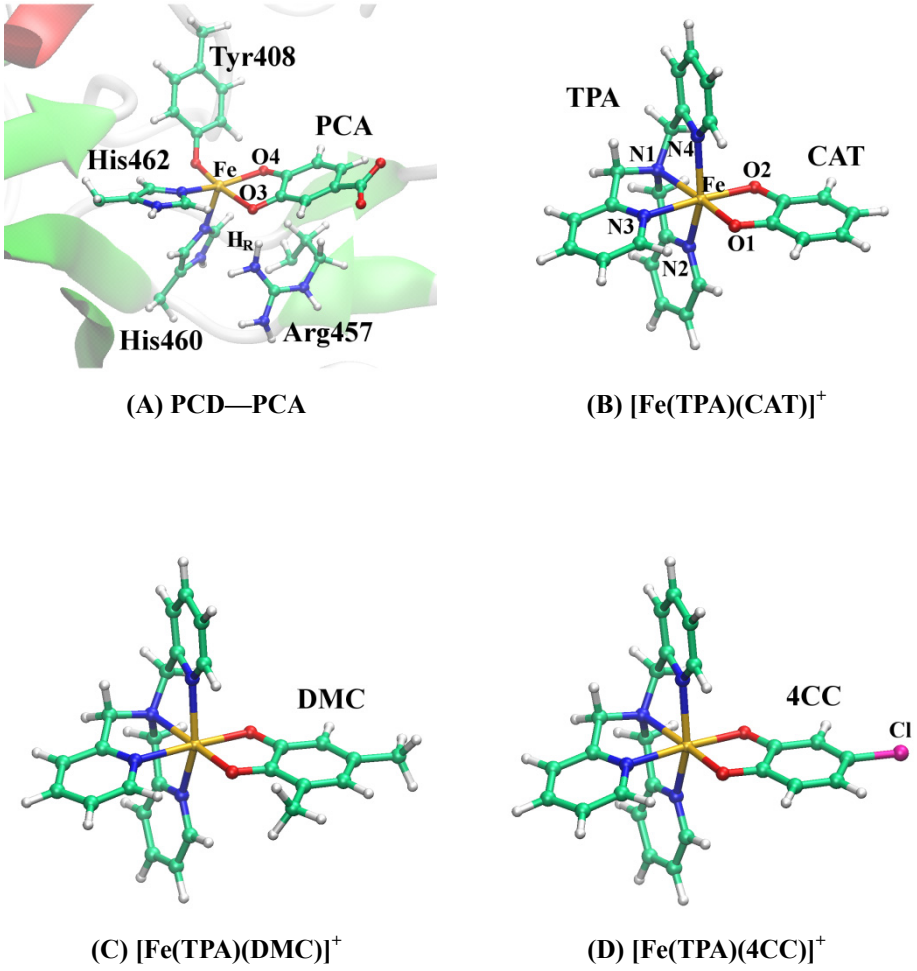
$[Fe(TPA)(R\text{-}CAT)]^+$ (TPA = tris(pyridylmethyl)amine and R-CAT = R-substituted catecholate; see Scheme 3) is one of the most important functional model complexes because of its high reactivity for dioxygen molecule.[11] Interestingly, the substitution on the catecholate ring induces considerably large shift of the LMCT absorption spectrum and significantly large change of reactivity for dioxygen molecule. Que and co-workers suggested that the LMCT excitation energy clearly relates to the reactivity for dioxygen molecule.[11a] Furthermore, Funabiki and co-workers proposed that there is the linear relationship between the LMCT excitation energy and the reactivity.[15] For instance, $[Fe(TPA)(CAT)]^+$ **2**, $[Fe(TPA)(DMC)]^+$ **3**, and $[Fe(TPA)(4CC)]^+$ **4** exhibit the LMCT absorption at 1.53, 1.39, and 1.58 eV, respectively and the rate constants for the dioxygen activation reaction are 1.3, 54, and 0.16 M^{-1} s^{-1}, respectively, in acetonitrile (MeCN) solution,[15] where CAT = catecholate (1,2-dihydroxybenzoate), DMC = 3,5-dimethylcatecholate, and 4CC = 4-chlorocatecholate. This is considered to be the experimental support that the iron(II)—SQ state contributes to the ground state. Thus, it is worthy of theoretically investigating the relation between the LMCT excitation energy and the reactivity for dioxygen molecule. The theoretical knowledge of the relation is useful to discuss the dioxygen activation by native iron(III)-dependent catechol dioxygenase, to understand the origin of high reactivity of native enzyme, and to design highly reactive bio-inspired oxygenation catalyst.

In this report, we theoretically investigate the LMCT excitation of the native iron(III)-dependent catechol dioxygenase and several model complexes to discuss the

electronic structures of both ground and excited states and the relation between the LMCT excitation energy and the reactivity for dioxygen molecule. If the resonance structure between iron(III)—CAT and iron(II)—SQ states exist, the iron(III)—CAT and the iron(II)—SQ states appear in close energy region. This means that the wave function of catecholatoiron(III) complex is represented by a mixture of these states. This is known as *quasi-degeneracy* problem. If so, catechol dioxygenase and its functional model complexes should be investigated with multi-reference electronic structure theory incorporating both non-dynamical and dynamical correlation effects. We employed here the multi-state (MS) CASPT2 method[16] because this method incorporates both non-dynamical and dynamical correlation effects.

6.2 Computational Details

We investigated here a substrate-bound complex of native iron(III)-dependent catechol dioxygenase (PCD—PCA) **1** and several functional model complexes, $[Fe(TPA)(CAT)]^+$ **2**, $[Fe(TPA)(DMC)]^+$ **3**, and $[Fe(TPA)(4CC)]^+$ **4**, as shown in Figure 1. In **1**, we added hydrogen atoms based on the experimental structure (PDB: 3PCA),[8] and determined the positions of added hydrogen atoms through partial molecular dynamics simulation in which hydrogen atoms were activated and the others were fixed to those of the experimental structure; see appendix in detail. Then, we performed geometry optimization by employing 2-layer ONIOM(QM:MM) method.[17] We involved the Fe, PCA, Tyr408, Arg457, His460, and His462 in a QM region and the others in a MM region, as shown in Scheme 4. We employed the density functional theory with B3LYP functional (DFT(B3LYP))[18] for the QM calculation and the AMBER96 force fields for the MM calculation. In the QM calculation, we employed Stuttgart/Dresden (311111/22111/411/1) ECP10MDF basis sets for Fe,[19]

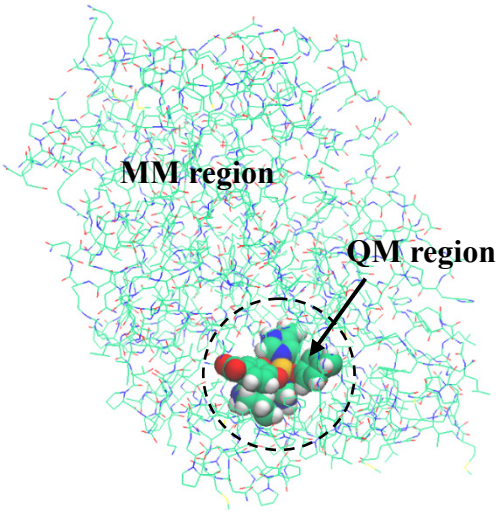


(A) PCD—PCA **(B) [Fe(TPA)(CAT)]$^+$**

(C) [Fe(TPA)(DMC)]$^+$ **(D) [Fe(TPA)(4CC)]$^+$**

Figure 1. (A) Optimized structure of native iron(III) catechol dioxygenase (PCD—PCA **1**) with the 2-layer ONIOM(DFT(B3LYP):AMBER) method and optimized structures of functional model complexes with the DFT(B3LYP) method; (B) [Fe(TPA)(CAT)]$^+$ **2**, (C) [Fe(TPA)(DMC)]$^+$ **3**, and (D) [Fe(TPA)(4CC)]$^+$ **4**. Important bond lengths (unit: Å): In **1**, Fe-O3 = 2.05, Fe-O4 = 2.37, Fe-O_{Tyr408} = 1.94, Fe-N_{His460} = 2.17, Fe-N_{His462} = 2.18, and O3-H_R = 1.68. In **2**, Fe-O1 = 2.01, Fe-O2 = 2.05, Fe-N1 = 2.29, Fe-N2 = 2.23, Fe-N3 = 2.19, and Fe-N4 = 2.23. In **3**, Fe-O1 = 2.01, Fe-O2 = 2.06, Fe-N1 = 2.28, Fe-N2 = 2.24, Fe-N3 = 2.19, and Fe-N4 = 2.24. In **4**, Fe-O1 = 2.01, Fe-O2 = 2.04, Fe-N1 = 2.28, Fe-N2 = 2.24, Fe-N3 = 2.18, and Fe-N4 = 2.23.

Scheme 4. Definition of QM and MM regions: Atoms described by sphere model are the QM region and the other atoms are the MM region.



6-31G* basis sets for H and C, and 6-31+G* basis sets for N and O. In the functional model complexes, **2**, **3**, and **4**, we performed the geometry optimization by the DFT(B3LYP) method with the same basis sets as those employed for the QM calculation of **1**.

To evaluate the excitation energy, we employed multi-state CASPT2 (MS-CASPT2) method,[16] because of the presences of large non-dynamical electron correlation in the Fe atom and quasi-degenerate wave function of catecholatoiron(III) complex. For Fe, we used the same basis sets as that used for DFT calculation, but atomic natural orbital (ANO) basis sets were used for the others; [4s2p/2s1p] for H in the catecholate moiety and [4s/2s] for other H, [10s6p4d/3s2p1d] for C in the catecholate moiety and [10s6p/3s2p] for other C, and [10s6p4d/4s3p1d] for N and O.[20] In **1**, we employed point charge (PC) approximation with AMBER96 charges to evaluate the protein electrostatic environment which is involved in the MM region. In the functional model complexes of **2**, **3**, and **4**, we employed polarized continuum model (PCM)[21] to evaluate solvation effect of acetonitrile (MeCN), which was used in the experimental study;[15] the dielectric constant of 36.64 (acetonitrile) was employed.

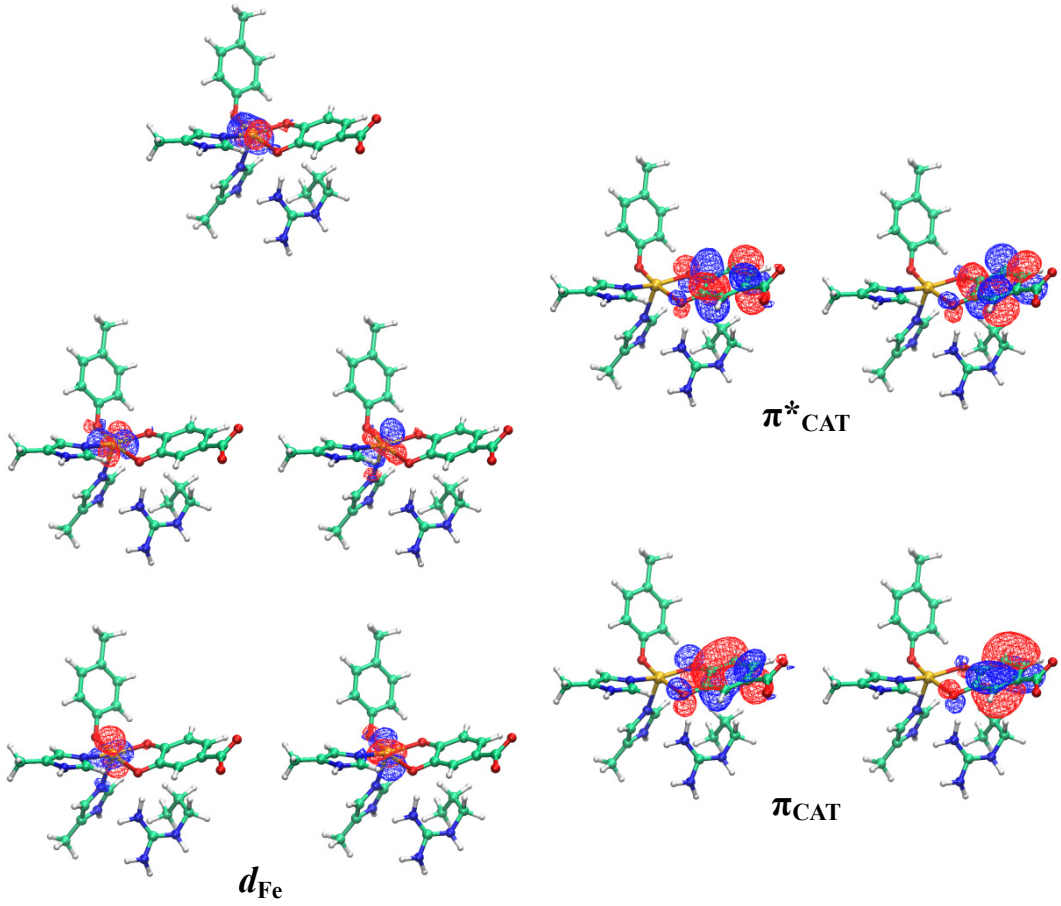
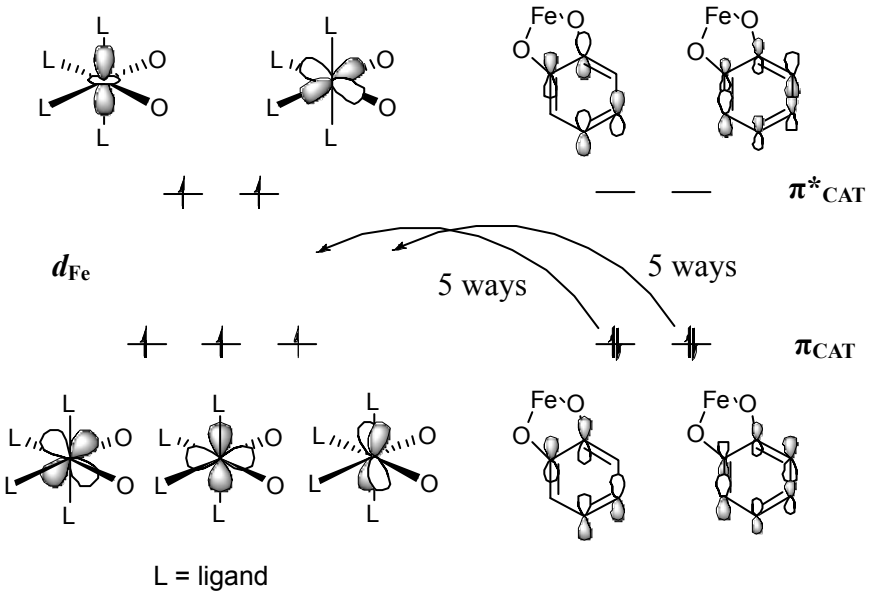


Figure 2. Averaged orbitals of **1** calculated by 11SA-CASSCF(9 in 9) method.

In MS-CASPT2 calculations, the CASSCF method[22] was employed to obtain a reference wave function. The active space of the CASSCF calculation includes nine electrons in such nine orbitals as five d orbitals of Fe and two sets of π and π* orbitals of catecholate moiety; see Figure 2 for the natural orbitals of **1**. This CASSCF calculation is represented by CASSCF(9 in 9) method, hereafter. In this case, one iron(III)—CAT configuration and ten iron(II)—SQ configurations participate in the ground and the LMCT excited states, as shown in Scheme 5. If the resonance structure between the iron(III)—CAT state and the iron(II)—SQ state exists, these states must be observed in a very close energy region. Thus, we carried out 11 state-averaged CASSCF(9 in 9) (11SA-CASSCF(9 in 9)) calculation to obtain the reference wave function, which contains one iron(III)—CAT state

Scheme 5



and ten iron(II)—SQ states. In MS-CASPT2 calculations, we employed the imaginary shift of 0.20 au to avoid intruder problem.[23]

All geometry optimizations were performed by Gaussian 03 (Revision C02) program package.[24] Molecular dynamics simulation was performed by AMBER 8 program package.[25] The MS-CASPT2 calculations were performed by Molcas 6.4 program package.[26]

6.3 Results and Discussions

6.3.1 Absorption spectra and electronic structures.

MS-CASPT2-calculated excitation energies are summarized in Table 1. These results agree well with the experimental values; RMS error is 0.09 eV for the first peak. In **1**, the excitation to the lowest energy excited state (2^1A) exhibits very small transition dipole. Because the intense absorption is experimentally observed as first peak, it is assigned as $X^1A \rightarrow 3^1A$ excitation, transition dipole of which is calculated to be considerably large. Also, the excitation to the 5^1A state of **1** exhibits somewhat large transition dipole, which

Table 1. The calculated LMCT excitation energies (ΔE in eV), the calculated transition intensities (*f* in au), and the experimental values of complexes **1**, **2**, **3**, and **4**. The excited states up to second peak of **1** and those up to first peaks of **2**, **3**, and **4** are summarized.

Cplx.	MS-CASPT2			TD-DFT			Exptl.
	state	ΔE / eV	*f* / au	state	ΔE / eV	*f* / au	ΔE / eV
	2^1A	1.40	0.01	2^1A	0.47	0.01	
				3^1A	0.63	0.00	
1	3^1A	**1.57**	**0.06**	4^1A	0.81	0.07	**1.65**[a]
	4^1A	1.79	0.01	5^1A	0.93	0.00	
	5^1A	2.02	0.09	6^1A	1.22	0.07	2.18[a]
	2^1A	1.08	0.00	2^1A	0.41	0.00	
2	3^1A	1.12	0.00	3^1A	0.46	0.00	
	4^1A	**1.45**	**0.17**	4^1A	1.41	0.12	**1.53**[b]
	2^1A	0.85	0.00	2^1A	0.35	0.00	
3	3^1A	0.86	0.00	3^1A	0.42	0.00	
	4^1A	**1.25**	**0.19**	4^1A	1.45	0.12	**1.39**[b]
	2^1A	1.21	0.00	2^1A	0.43	0.00	
4	3^1A	1.24	0.00	3^1A	0.48	0.00	
	4^1A	**1.54**	**0.17**	4^1A	1.36	0.15	**1.58**[b]

a) reference 11.

b) reference 14.

corresponds to the second peak experimentally observed.[13] These two excitation energies are calculated to be 1.57 and 2.02 eV, respectively, by the MS-CASPT2 method, which both agree well with their experimental values of 1.65 and 2.18 eV. In **2**, **3**, and **4**, transition dipoles of $X^1A \rightarrow 2^1A$ and $X^1A \rightarrow 3^1A$ excitations were calculated to be almost zero. Thus, the first peaks of **2**, **3**, and **4** are assigned as $X^1A \rightarrow 4^1A$ excitation. The calculated excitation energy increases in the order **3** < **2** < **4** < **1**. This increasing order agrees with that of the experimentally observed first peak.[13, 15] These results indicate that the MS-CASPT2 method presents reliable results of the electronic structures of **1** to **4**. We also performed time-dependent DFT (TD-DFT)[27] calculations to evaluate the excitation energies of the

catecholatoiron(III) complexes. As summarized in Table 1, however, the TD-DFT-calculated results do not agree with the experimental values even qualitatively; for instance, the TD-DFT-calculated excitation energy increases in the order **1** < **4** < **2** < **3**, which is reverse to the experimental trend. These results indicate that the multi-reference electronic structure theory including dynamical and non-dynamical electron correlation effects is crucial to discuss the ground and excited states of these catecholatoiron(III) complexes. We discussed LMCT excitation energies and electronic structures based on the MS-CASPT2 wave function, hereafter.

Interestingly, "*dark states*", which are excited states bearing either very small or no transition dipole, are calculated to be in lower energy region than the first peak, as mentioned above (see also Table 1). The excitation energies of the lowest energy excited states of **1**, **2**, **3**, and **4** are calculated to be 1.40, 1.08, 0.85, and 1.21 eV, respectively, and their transition dipoles are calculated to be 0.01 au at most. If the excitation energy is very small, those dark states should be considered in thermal equilibrium. However, these dark states still exist at high energies compared to the thermal energy, not to contribute to the ground state (maximum of 4×10^{-15}).

To discuss electronic structures of the ground and excited states, we focused on 11 electron configurations described in Scheme 5. This is because these 11 configurations are of considerable importance in the ground and the LMCT excited states of catecholatoiron(III) complex. We evaluated a weight of an iron(III)—CAT configuration $w_{\mathrm{Fe(III)}}$ and a weight of iron(II)—SQ configuration $w_{\mathrm{Fe(II)}}$ which are the contribution of the iron(III)—CAT state and that of the sum of the iron(II)—SQ states, respectively. In the ground states, the $w_{\mathrm{Fe(III)}}$ is 0.80, 0.74, 0.68, and 0.76 for **1**, **2**, **3**, and **4**, respectively, and the $w_{\mathrm{Fe(II)}}$ is 0.13, 0.19, 0.25, and 0.17 for **1**, **2**, **3**, and **4**, respectively. These results clearly indicate that the ground states of

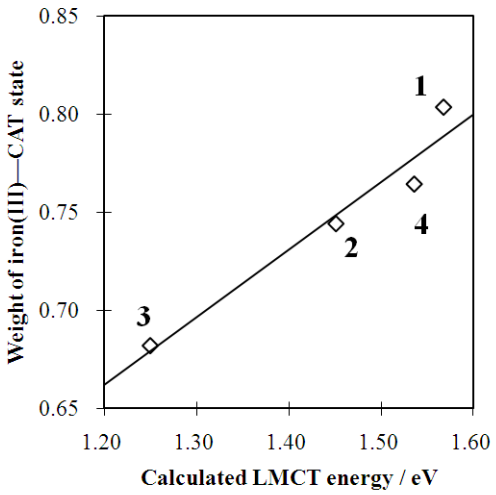


Figure 3. Relationship between the MS-CASPT2-calculated LMCT energies and weights of iron(III), $w_{Fe(III)}$, in the ground states of complexes **1**, **2**, **3**, and **4**.

these catecholatoiron(III) complexes mainly consist of the iron(III)—CAT state (68 ~ 80%), and that the iron(II)—SQ states is moderately contribute to the ground state; their contributions are 13 ~ 25%. In other words, the ground state is represented by the wave function of multi-reference character, which corresponds to more chemical representation of "*resonance structure*". The contribution of the iron(II)—SQ states is considerably larger than the thermal contribution of the dark states. This is consistent with previous experimental study.[14]

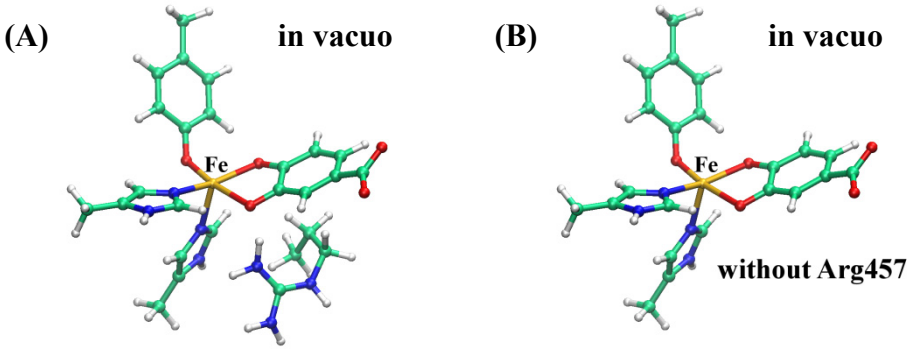
Interestingly, the weight of iron(III)—CAT $w_{Fe(III)}$ is proportional to the calculated and the observed LMCT excitation energies, as shown in Figure 3, indicating that the LMCT excitation energy clearly relates to the ground state electronic structure; the decrease of iron(III)—CAT character and the increase of iron(II)—SQ character in the ground state leads to red shift of the LMCT excitation energy. Also, the $w_{Fe(III)}$ relates to the reactivity for dioxygen molecule in functional models; for instance, the reactivity increases **4** < **2** < **3** which is the same as the decreasing order of iron(III)—CAT character in the ground state.

However, this relation is not observed in **1**; though the $w_{Fe(III)}$ of **1** was calculated to be the largest among the complexes examined here, **1** is the most reactive for dioxygen molecule. This indicates that other factors are responsible for the reactivity of the native enzyme **1**, which will be discussed below.

6.3.2 Environmental effects.

Here, we wish to make comparison among the catecholatoiron(III) complex in vacuo, protein, and solvent. First, we investigated the active site of native enzyme **1** and the functional model complex **2** without any environmental effects; **1** and **2** in vacuo are named $\mathbf{1}_{vac}$ and $\mathbf{2}_{vac}$, respectively. In $\mathbf{1}_{vac}$, only the QM region, [Fe(MePhO)(MeIm)$_2$(PrG)(PCA)], was considered, where *p*-MePhO (*p*-methylphenoxide), MeIm (3*H*-4-methylimidazole), and PrG (guanidinopropane) are employed as models of tyrosine, δ-protonated histidine, arginine, respectively, as shown in Figure 4(A).

The excitation energies of $\mathbf{1}_{vac}$ and $\mathbf{2}_{vac}$ are calculated to be 1.31 and 1.22 eV, respectively, which are much smaller than those of **1** and **2** (Table 2). Interestingly, the $w_{Fe(III)}$ value is 0.19 and 0.35 for $\mathbf{1}_{vac}$ and $\mathbf{2}_{vac}$, respectively, which are much smaller than those of **1** and **2** (Table 2). These results indicate that the ground state mainly consists of the iron(II)—SQ state in vacuo, which are completely different from those of **1** in protein and **2** in solution. Also, the absorptions of $\mathbf{1}_{vac}$ and $\mathbf{2}_{vac}$ are assigned to be metal-to-ligand charge transfer (MLCT) excitations from the iron(II) center to the semiquinonate moiety. Thus, it is concluded that the environmental effects are of considerable importance and the electronic structure of catecholatoiron(III) complex completely changes by the environment. Here, we investigate two important factors about protein environmental effects; one is electrostatic interaction by protein and another is hydrogen bonding with amino acid residue.



[Fe(MePhO)(MeIm)$_2$(PrG)(PCA)] [Fe(MePhO)(MeIm)$_2$(PCA)]$^-$

Figure 4. Model systems for the active site in vacuo. (A) Active site model which considered as the QM region in protein phase calculations (see also computational details) and (B) active site model without Arg457 from (A), where MePhO = *p*-methylphenoxide (tyrosine model), MeIm = 3*H*-4-methylimidazole (δ-protonated histidine model), and PrG = guanidinopropane (arginine model). Both geometries were constrained to be the same as that optimized by 2-layer ONIOM(DFT(B3LYP):AMBER) method.

Table 2. Environmental effects on the LMCT excitation energy and weight of iron(III)—CAT configuration $w_{Fe(III)}$.

native enzyme		ΔE / eV	$w_{Fe(III)}$
in vacuo	**1'**$_{vac}$	1.71	0.05
+ Arg457	**1**$_{vac}$	1.31	0.19
+ electrostatic interaction	**1**	1.57	0.80

[Fe(TPA)(CAT)]$^+$		ΔE / eV	$w_{Fe(III)}$
in vacuo	**2**$_{vac}$	1.22	0.35
+ solvation (MeCN)	**2**	1.45	0.74

6.3.3 Protein environment (1): electrostatic interaction with surrounding region.

The electrostatic interaction between the active site and each amino acid residue is approximately represented by classical electrostatic interaction energy E_{ES} defined by eq. (2);

$$\begin{aligned} E_{ES} &= \sum_{i}^{i\in MM} E_{ES,i} \\ &= \sum_{i}^{i\in MM} \left(\sum_{A}^{A\in i} \sum_{B}^{B\in QM} \frac{Q_A Q_B}{r_{AB}} \right) \end{aligned} \qquad (2),$$

where index *i* denotes each residue and Q_A is charge on the atom A. We further introduced the energy difference, ΔE_{ES}, between the iron(III)—CAT and the iron(II)—SQ states defined by

$$\begin{aligned} \Delta E_{ES} &= \sum_{i}^{i\in MM} \Delta E_{ES,i} \\ &= \sum_{i}^{i\in MM} \left(E_{ES,i}^{iron(II)} - E_{ES,i}^{iron(III)} \right) \\ &= \sum_{i}^{i\in MM} \left(\sum_{A}^{A\in i} \sum_{B}^{B\in QM} \frac{Q_A Q_B^{iron(II)}}{r_{AB}} - \sum_{A}^{A\in i} \sum_{B}^{B\in QM} \frac{Q_A Q_B^{iron(III)}}{r_{AB}} \right) \\ &= \sum_{i}^{i\in MM} \left(\sum_{A}^{A\in i} \sum_{B}^{B\in QM} \frac{Q_A (Q_B^{iron(II)} - Q_B^{iron(III)})}{r_{AB}} \right) \end{aligned} \qquad (3),$$

where the negative ΔE_{ES} value represents that the electrostatic interaction more stabilizes the iron(III)—CAT state than the iron(II)—SQ state.

We employed the AMBER96 charges for Q_A and the Mulliken charges of the iron(III)—CAT state ($Q_B^{Fe(III)}$) and the iron(II)—SQ state ($Q_B^{Fe(II)}$), where $Q_B^{Fe(III)}$ and $Q_B^{Fe(II)}$ were calculated by 11SA-CASSCF(9 in 9) method. As shown in Figure 5(A), Arg133 gives positively large ΔE_{ES} value and Asp65 gives negatively large ΔE_{ES} value. This is because positively charged Arg133 and negatively charged Asp65 are close to the deprotonated carboxyl group of the iron(III)—CAT state (Figure 5(B)). The protein electrostatic interaction totally stabilizes the iron(III)—CAT state. In other words, Arg133 plays key role

to change the ground state to the iron(III)—CAT state from the iron(II)—SQ state.

6.3.4 Protein environment (2): hydrogen bond with Arg457.

The hydrogen bonding interaction between PCA and Arg457 is expected to be one of the strongest contributions to the active site because the O3-H_R distance is considerably short (1.68 Å). To investigate how much this hydrogen bond contributes to the electronic structure, we calculated the model complex $[Fe(MePhO)(MeIm)_2(PCA)]^-$ in vacuo (as represented by **1'**$_{vac}$), in which the Arg457 was omitted from **1**$_{vac}$, as shown in Figure 5(B). The excitation energy and $w_{Fe(III)}$ were calculated to be 1.71 eV and 0.05, respectively (Table 2). Interestingly, the $w_{Fe(III)}$ of **1'**$_{vac}$ is considerably smaller than that (0.19) of **1**$_{vac}$, which indicates that the ground state is almost the iron(II)—SQ state. It should be noted that

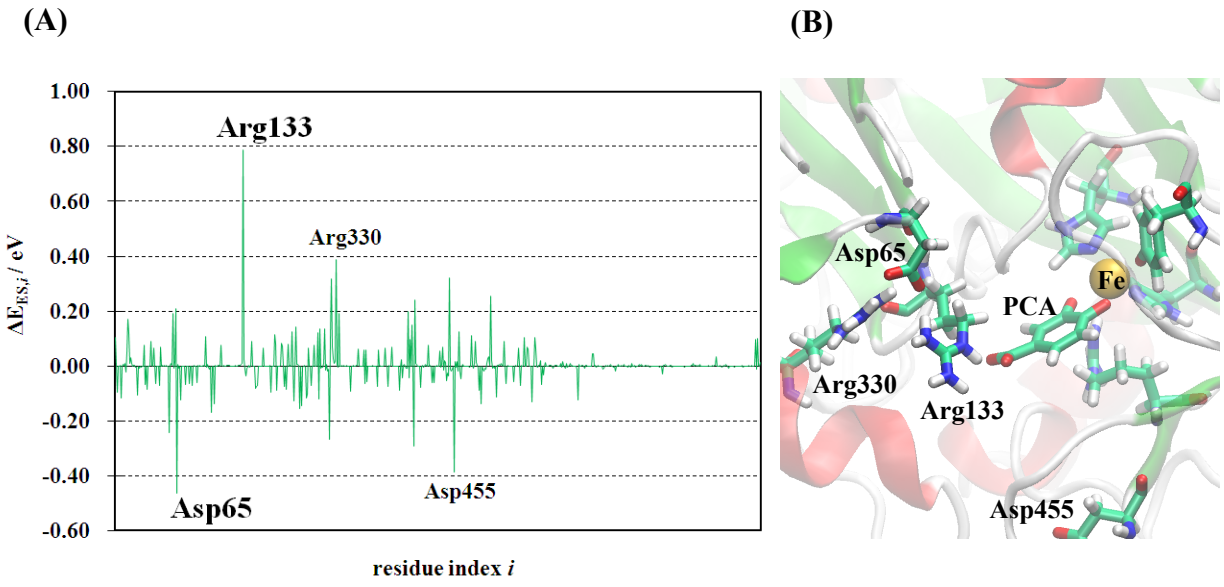


Figure 5. (A) Difference electrostatic interaction energy between the iron(III)—CAT state and the iron(II)—SQ state for each residue *i* ($\Delta E_{ES,i}$ as defined by Eq.(3)), in which positive sign means the residue *i* favors the iron(III)—CAT state and negative sign means the residue *i* favors the iron(II)—SQ state. (B) Residues which give remarkable contribution in $\Delta E_{ES,i}$.

though the excitation energy is not different very much from that (1.57 eV) of **1**, this excitation is assigned to be the MLCT excitation which is completely different from that of **1**, as mentioned above. This result indicates that the Arg457 stabilizes the π orbital energy of catecholate moiety through the hydrogen bonding interaction to stabilize the iron(III)—CAT state.

6.3.5 Solvation effect.

In the functional model complexes, we employed the PCM method to consider the solvation effect of MeCN.[15] As discussed above, the $w_{Fe(III)}$ is considerably larger in **2** than in $\mathbf{2}_{vac}$. Because MeCN is non-protic polar solvent, the solvation effect can be simply understood by considering the dipole-dipole interaction between catecholatoiron(III) complex and solvent. From the 11SA-CASSCF wave function of $\mathbf{1}_{vac}$, the dipole moment of iron(III)—CAT state was calculated to be 18.7 D but those of iron(II)—SQ states were calculated to be 4.3-4.6 D. This result suggests that the dipole moment of the iron(III)—CAT state is approximately described by the interaction between +3 and −2 centers but that of the iron(II)—SQ states are approximately described by the interaction between +2 and −1 centers. Thus, the polar solvent considerably stabilizes the iron(III)—CAT state to lead to the increase of $w_{Fe(III)}$ in the ground state and blue shift of the LMCT excitation energy.

6.3.6 Relation between the electronic structure and the reactivity.

Based on the CASPT2 computational results, we previously reported that the direct charge transfer from the catecholate moiety to dioxygen molecule occurs in dioxygen activation process.[28] The iron(III) center plays important role to induce the charge transfer by lowering the energy level of π* orbital of the dioxygen molecule.

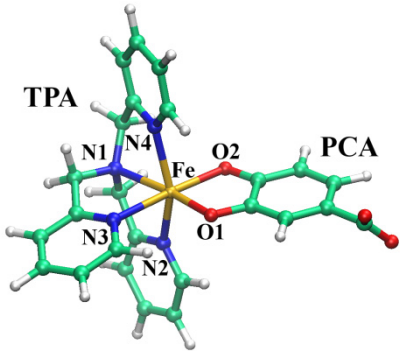


Figure 6. Optimized structure of [Fe(TPA)(PCA)] **5** with DFT(B3LYP) method. Important bond lengths (unit: Å): Fe-O1 = 1.96, Fe-O2 = 2.09, Fe-N1 = 2.37, Fe-N2 = 2.23, Fe-N3 = 2.21, and Fe-N4 = 2.22.

We focus on the π orbital energy of catecholate because the charge transfer (CT) from the catecholate moiety to the dioxygen molecule becomes stronger as the π orbital energy becomes higher. However, orbital energy cannot be defined in CASSCF and CASPT2 wave functions. We introduced here the ionization potential of catecholate moiety (IP_{CAT}) instead of the orbital energy, because the IP_{CAT} directly relates to the π orbital energy of catecholate moiety.[29] The IP_{CAT} was defined by eq. (1):

$$IP_{CAT} = E(\text{iron(III)}-\text{CAT}) - E(\text{iron(III)}-\text{SQ}) \qquad (1).$$

E(iron(III)—CAT) and *E*(iron(III)—SQ) are evaluated here with the CASPT2(9 in 9) and CASPT(8 in 9) methods, respectively. The calculated-IP_{CAT} decreases in the order **4** (6.04eV) > **2** (5.99 eV) > **3** (5.71 eV), where in parentheses are IP_{CAT} values. This trend agrees with the experimental facts that the reactivity increases in the order **4** < **2** < **3**[15] and the LMCT excitation energy decreases in the order **4** > **2** > **3**. These results are consistent with the experimental expectation that the reactivity for dioxygen molecule increases as the LMCT excitation energy decreases.

However, the IP_{CAT} of **1** was calculated to be almost the same as that of

[Fe(TPA)(DMC)]$^+$ **3**, though its reactivity is the highest and its LMCT excitation energy of **1** is calculated to be considerably higher than that of **3**. These results indicate that not only the IP_{CAT} but also other factors contribute to its reactivity and the LMCT excitation energy. Because not CAT but PCA is a substrate in the native enzyme, we employed the model complex [Fe(TPA)(PCA)] **5** (Figure 6). The IP_{CAT} of **5** was calculated to be 5.82 eV, which is moderately larger than that (5.72 eV) of the PCD—PCA complex **1** but moderately smaller than that (5.99 eV) of [Fe(TPA)(CAT)]$^+$ **2**. This result indicates that the PCA is moderately more reactive than the non-substituted catecholate (CAT) and the dioxygen activation easier occurs in the native catechol dioxygenase than in [Fe(TPA)(CAT)]$^+$. However, it is not enough to explain the highest reactivity of **1**. The native enzyme has 5-coordinate iron center, in which dioxygen molecule can easily approach the sixth site to form 6-coordinate iron intermediate. While [Fe(TPA)(CAT)]$^+$ has already 6-coordinate iron center, which requires significant structural change in dioxygen binding. From experimental study, it has been mentioned that the rate-determining-step (RDS) of native enzyme is O-O bond cleavage process but that of 6-coordinate model complex is O_2 binding process.[30] This experimental fact and our computational results suggest that unsaturated 5-coordinate iron(III) center is crucial to the high reactivity of native enzyme.

6.4 Conclusions

We reported here the MS-CASPT2 study on the ground and excited states of such catecholatoiron(III) complexes as native iron(III)-dependent catechol dioxygenase **1** and its functional model complexes **2**, **3**, and **4**. Our MS-CASPT2-calculated results quantitatively reproduced the experimental absorption spectra of catecholatoiron(III) complexes. Thus, MS-CASPT2 method is useful to the quantitative discussion about various electronic

structures of both ground and excited states.

The ground state of catecholatoiron(III) complex mainly consists of the iron(III)—catecholate form and moderately the iron(II)—semiquinonate form in the native enzyme 1 and model complexes **2** to **4**; the contribution of the iron(II)—semiquinonate form is at most 30%. Environmental effects are of considerable importance because the ground state electronic structure remarkably changes by the environment. Because the ground state of catecholatoiron(III) complex in **1** and **2** to **4** is calculated to mainly consist of the iron(II)—semiquinonate form in vacuo. In the PCD, Arg457 and Arg133 play important roles to stabilize the iron(III)—catecholate form. Arg457 directly interacts with catecholate-oxygen to form the strong hydrogen bond and to stabilize the iron(III)—catecholate form. Arg133 provides positively large electrostatic field on the catecholate, which also stabilizes the iron(III)—catecholate form. In **2** to **4**, the polar solvent plays crucial role, as follows: Because the dipole moment of iron(III)—catecholate form is considerably larger than that of iron(II)—semiquinonate form, the iron(III)—catecholate form is peculiarly stabilized in polar solvent such as acetonitrile (MeCN).

The LMCT excitation energy clearly reflects in the ground state electronic structure; the increase of iron(II)—semiquinonate form causes the red-shift of the LMCT excitation energy. The reactivity with dioxygen molecule relates to the energy level of catecholate π orbital.[28] Here, we introduced the ionization potential of the catecholate moiety (IP_{CAT}), which is the key property related to the reactivity for dioxygen molecule. The reactivity of functional models clearly relates to the IP_{CAT}. However, the reactivity of native iron(III)-dependent catechol dioxygenase cannot be explained only by the IP_{CAT} value, indicating that the unsaturated 5-coordinate iron(III) center found in the native enzyme plays essential role for the high reactivity.

6.5 Appendix

MD Annealing Calculation

To obtain reasonable initial structure for the ONIOM(QM:MM) geometry optimization, we performed a partial molecular dynamics simulation for H atoms. First, we added H atoms to the X-ray crystal structure of PCD—PCA complex (PDB-ID: 3PCA). Second, we carried out the energy minimization for H atoms in which all heavy atoms were constrained to the experimental structure (heavy atoms were also constrained in following calculations). Third, we performed 1 ns simulation for increasing temperature to 300 K and followed by 9 ns simulation for the sampling at 300 K. Finally, we picked up 10 structures from the sampling and we carried out the energy minimization again for each structure. We employed the structure which gives the lowest potential energy among 10 annealed structures as an initial structure of the ONIOM(QM:MM) geometry optimization. Energy minimizations and molecular dynamics simulations were carried out by AMBER 8 program package with AMBER parm96 force fields. Despite only the positions of H atoms were relaxed, the annealed structure was calculated to be more stable than the initially minimized structure by 43.7 kcal/mol in the potential energy. This is because the hydrogen bond network will be consistent. Thus, the protein structure should be carefully constructed before QM calculations.

References

(1) Hayaishi, O.; Hashimoto, K. *J. Biochem.* **1950**, *37*, 371-374.

(2) Costas, M.; Mehn, M. P.; Jensen, M. P.; Que, L., Jr. *Chem. Rev.* **2004**, *104*, 939-986.

(3) Gibson, D. T. In *Microbial Degradation of Organic Compounds*; Marcel Dekker: New York, **1984**.

(4) Lipscomb, J. D.; Orville, A. M. In *Metal Ions in Biological Systems*; Sigel, H., Sigel, A., Eds.; Marcel Dekker: New York, **1992**; Vol. 28, pp 243-298.

(5) Funabiki, T. In *Oxygenases and Model Systems*; Funabiki, T., Ed.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1997; pp19-104.

(6) Bugg, T. D. H. *Tetrahedron* **2003**, *59*, 7075-7101.

(7) (a) Ohlendorf, D. H.; Lipscomb, J. D.; Webber, P. C. *Nature* **1988**, *336*, 403-405.; (b) Ohlendorf, D. H.; Orville, A. M.; Lipscomb, J. D. *J. Mol. Biol.* **1994**, *244*, 429-440.

(8) Orville, A. M.; Lipscomb, J. D.; Ohlendorf, D. H. *Biochemistry* **1997**, *36*, 10052-10066.

(9) Note that we used the expressions of iron(III)—CAT and iron(II)—SQ to represent an electronic state or an electron configuration; it was distinguished from a complex name i.e., catecholatoiron(III).

(10) Bull, C.; Ballou, D.; Otsuka, S. *J. Biol. Chem.* **1981**, *256*, 12681-12686.

(11) (a) Jang, H. G.; Cox, D. D.; Que, L., Jr. *J. Am. Chem. Soc.* **1991**, *113*, 9200-9204.; (b) Raffard, N.; Carina, R.; Simaan, A. J.; Sainton, J.; Rivière, E.; Tchertanov, L.; Bourcier, S.; Bouchoux, G.; Delroisse, M.; Banse, F.; Girerd, J.-J. *Eur. J. Inorg. Chem.* **2001**, 2249-2254.

(12) Que, L., Jr.; Kolanczyk, R. C.; White, L. S. *J. Am. Chem. Soc.* **1987**, *109*, 5373.

(13) Pau, M. Y. M.; Davis, M. I.; Orville, A. M.; Lipscomb, J. D.; Solomon, E. I. *J. Am. Chem. Soc.* **2007**, *129*, 1944-1958.

(14) Funabiki, T.; Fukui, A.; Hitomi, Y.; Higuchi, M.; Yamamoto, T.; Tanaka, T.; Tani, F.; Naruta, Y. J. Inorg. Biochem. **2002**, *91*, 151-158.

(15) Hitomi, Y.; Yoshida, M.; Higuchi, M.; Minami, H.; Tanaka, T.; Funabiki, T. *J. Inorg. Biochem.* **2005**, *99*, 755-763.

(16) Finley, J.; Malmqvist, P.-Å.; Roos, B. O.; Serrano-Andrés, L. *Chem. Phys. Lett.* **1998**, *288*, 299-306.

(17) Dapprich, S.; Komaromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *J. Mol. Struct. Theochem.* **1999**, *461-462*, 1-21.

(18) (a) Lee, C.; Yang, W.; Parr, R.G. *Phis. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785-789.; (b) Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. *Chem. Phys. Lett.* **1989**, *157*, 200-206.; (c) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648-5652.

(19) Dolg, M.; Wedig, U.; Stoll, H.; Preuss, H. *J. Chem. Phys.* **1987**, *86*, 866-872.

(20) Pierloot, K.; Dumez, B.; Widmark, P.-O.; Roos, B. O. *Theor. Chim. Acta.* **1995**, *90*, 87-114.

(21) (a) Cancès, M. T.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032-3041.; (b) Cossi, M.; Barone, V.; Mennucci, B.; Tomasi, J. *Chem. Phys. Lett.* **1998**, *286*, 253-260.; (c) Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *106*, 5151-5158.

(22) Roos, B. O.; Taylor, P. R. *Chem. Phys.* **1980**, *48*, 157-173.; (b) Roos, B. O. In *Ab Initio Methods in Quantum Chemistry II*; Lawley, K. P., Ed.; Wiley: New York, **1987**, p.399.

(23) Forsberg, N.; Malmqvist, P.-Å. *Chem. Phys. Lett.* **1997**, *274*, 196-204.

(24) Frisch, M. J.; et al. *Gaussian 03*, Revision C.02; Gaussian Inc.; Wallingford, CT, **2004**.

(25) Case, D. A.; Darden, T. A.; Cheatham, III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Gui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J.

W.; Ross, W. S.; Kollman, P. A. AMBER 8, **2004**, University of California, San Francisco.

(26) Karlström, G.; Lindh, R.; Malmqvist, P.-Å.; Roos, B. O.; Ryde, U.; Veryazov, V.; Widmark, P.-O.; Cossi, M.; Schimmelpfennig, B.; Neogrady, P.; Seijo, L. *Comput. Mater. Sci.* **2003**, *28*, 222-239.

(27) Bauernschmitt, R.; Ahlrichs, R. *Chem. Phys. Lett.* **1996**, *256*, 454-464.

(28) Nakatani, N.; Nakao, Y.; Sato, H.; Sakaki, S. *J. Phys. Chem. B.* **2009**, *113*, 4826-4836.

(29) In Koopmans theorem, orbital energy corresponds to the ionization potential (for occupied orbitals) or the electron affinity (for unoccupied orbitals) in molecular orbital theories. However, because the CASSCF/CASPT2 wave function does not have an orbital energy picture we considered here the ionization potential and the electron affinity, which are effectively the same as the discussions with orbital energy, furthermore, these are much accurate discussions rather than using orbital energy.

(30) (a) Walsh, T. A.; Ballou, D. P.; Mayer, R.; Que, L., Jr. *J. Biol. Chem.* **1983**, *258*, 14422-14427. (b) Merkel, M.; Pascaly, M.; Krebs, B.; Astner, J.; Foxon, S. P.; Schindler, S. *Inorg. Chem.* **2005**, *44*, 7582-7589.

Chapter 7

General Conclusion

In this thesis, the author theoretically investigated the biological systems and biological phenomena, in particular, firefly bioluminescence and dioxygen activation by non-heme iron enzyme in which the complicated electronic structure plays essential roles for protein function. The important findings presented in this thesis provide us with the fundamental knowledge of biological systems and biological phenomena. The author believes that such knowledge is indispensable to make further development in the chemistry of biological systems.

In chapter 1, the yellow-green bioluminescence from firefly oxyluciferin was theoretically investigated with SAC/SAC-CI method. Firefly oxyluciferin undergoes keto-enol tautomerization. The keto-oxyluciferin exhibits the red chemiluminescence and the enol-oxyluciferin exhibits the yellow-green chemiluminescence. Based on the similarity between colors of chemiluminescence and bioluminescence, the yellow-green bioluminescence of firefly luciferase has been understood for a long time to be the emission from the enol-oxyluciferin. However, a recent experiment proposed that the oxyluciferin keeps the keto form and it participate in yellow-green emission in luciferase. The SAC/SAC-CI-calculated emission energy is close to the experimental value. It is clearly shown by this study well supportive that the keto-form participates in the yellow-green bioluminescence of firefly luciferase. The twisted intramolecular charge transfer (TICT) state was also examined because previous study proposed that the TICT excited state of keto-oxyluciferin exhibits the yellow-green emission. However, the TICT excited state is

clearly ruled out by the SAC/SAC-CI computational study, because the TICT excited state is meta stable and it easily relaxes to the co-planer structure.

In chapter 2, the color-tuning mechanism of firefly luciferase was theoretically proposed by the SAC/SAC-CI study. Recently, firefly bioluminescence is used for ATP assays in brain cell. Because red light is less scattered by water molecule in cell, red-colored bioluminescence by luciferase mutant is very useful. The author analyzed the electrostatic interactions between the firefly oxyluciferin and the surrounding proteins and identified what amino acids play crucial roles for the color-tuning of emission spectrum. The SAC/SAC-CI computational results indicate that the neighboring residues, in particular, Arg218 and the phosphate group of AMP play important roles to exhibit the yellow-green luminescence from the keto-form oxyluciferin. However, such residues were experimentally expected to drastically change the active site to induce considerable decrease of quantum yield. Interestingly, the analysis of electrostatic interaction suggests that there are several residues which do not directly interact with the active site but can also play important roles for the color-tuning; they are Arg223, Glu344, and Asp422. From these results, the author wishes to propose the Arg223Ala/Glu344Ala/Asp422Ala mutant of *Photinus pyralis* luciferase a red-color firefly luciferase, because these mutations artificially change the luminescence color through the electrostatic effect on the luciferase protein. The theoretical calculations of these mutant shows that the emission energy was estimated to be 2.05 eV (602 nm, red-dish-orange), which is 0.18 eV lower than that of the wild type (2.23eV, 557 nm, yellow-green).

In chapter 3, the author theoretically investigated the active center of iron-sulfur-cluster-free hydrogenase (Hmd) with the DFT method. Interestingly, Hmd contains mononuclear low-spin iron center, which is considerably different from the binuclear

active center of typical [NiFe]-hydrogenase. However, the geometry and the electronic structure of the active site of Hmd are still not clear. Based on DFT-calculated CO stretching frequency, it is clearly concluded that the active site contains low-spin iron(II) center coordinated with two COs, cysteine, either acylpyridone and water or acylpyridinol and hydroxide. Computational results of geometrical features and bonding nature are consistent with experimental understanding.

In chapter 4, dioxygen binding process of non-heme iron(III) center in intra-diol catechol dioxygenase was theoretically investigated. The author employed here the CASSCF/CASPT2 method to incorporate multi-configurational character participating in Fe-O_2 interaction. In this process, two alternative mechanisms were experimentally proposed; one is called "Oxygen Activation" and the other is called "Substrate Activation". Our CASSCF/CASPT2-calculated results support the "Oxygen Activation". Potential energy curves and electronic structure evaluated with SA(State-Averaged)-CASSCF/CASPT2 method indicate that the charge transfer directly occurs from the catecholate moiety to the dioxygen moiety in the O_2 binding process, to produce an η^1-end-on type iron(III)-superoxo complex. This is the key step of the dioxygen activation. Interestingly, the iron center always keeps high spin d^5 character during the O_2 binding process, indicating that the iron(III) center does not receive charge transfer from the catecholate moiety. However, this does not mean that the iron(III) center does not play any role in the dioxygen activation. The important role of the iron(III) center is to adjust the π^* orbital energy level of dioxygen molecule to induce the charge transfer from the catecholate moiety to the dioxygen moiety. Besides the η^1-end-on iron(III)-superoxo complex, η^2-side-on type iron(III)-superoxo complex is also optimized. This species is more stable than the η^1-end-on type iron(III)-superoxo complex, suggesting that this is considered as a stable isomer in the early

stage of the catalytic cycle.

In chapter 5, the author theoretically investigated the ligand-to-metal charge-transfer (LMCT) excitation of the native iron(III) catechol dioxygenase and its functional model complexes. Based on the idea that the charge transfer occurs from the catecholate moiety to the iron(III) center to afford an iron(II)—semiquinonate species which is expected to be reactive for dioxygen activation, the reactivity of catechol dioxygenase models is often discussed in terms of the LMCT absorption energy. The ground and excited states of these complexes were investigated with MS-CASPT2 method. The present MS-CASPT2 calculations quantitatively reproduced the experimental absorption peak. The MS-CASPT2 wave function indicates that the ground state of catecholatoiron(III) complex mainly consists of the iron(III)—catecholate electron configuration and moderately of the iron(II)—semiquinonate electron configuration at most 30%. The protein environmental effects are considerably important, as follows: The ground state electronic structure mainly consists of iron(II)—semiquinonate electron configuration in vacuo, while it mainly consists of iron(III)—catecholate electron configuration in the protein environment. The LMCT excitation energy clearly relates to the weight of the iron(II)—semiquinonate configuration in the ground state, but it does not directly relate to the reactivity for dioxygen activation. The reactivity directly relates to the catecholate π orbital energy, because the π orbital participates in the CT to the dioxygen molecule. The energy level of the π orbital depends on both substituent on catecholate ring and ligand of iron(III) complex.

Nowadays, electronic structure theory, in particular, density functional theory, has been succeeded to understand geometry, bonding nature, reaction mechanism, and physicochemical property of various atoms and molecules. However, theoretical investigations of biological

systems still have lots of problems because biological systems are generally very large and complicated. Particularly, in the photobiological systems and the metalloenzymes, the complicated electronic structures play crucial roles for their functions. They require the use of the accurate electronic structure theories in theoretical investigation of such biological phenomena. However, because of their huge computational costs the accurate electronic structure theories have been little employed in theoretical investigation of biological systems. The author started the theoretical studies of biological systems with such accurate electronic structure theories as SAC/SAC-CI and CASSCF/CASPT2. His studies showed that the use of the accurate electronic structure theories is indispensable to understand well the biological systems and biological phenomena. The author believes that such electronic structure theories must be much more applied to theoretical investigations of biological systems.

List of Publications

Publications included in this thesis

Chapter 2

"Red Light in Chemiluminescence and Yellow-Green Light in Bioluminescence: Color-Tuning Mechanism of Firefly, *Photinus pyralis*, Studied by the Symmetry-Adapted Cluster-Configuration Interaction Method"

Naoki Nakatani, Jun-ya Hasegawa, and Hiroshi Nakatsuji

J. Am. Chem. Soc. **2007**, *129*, 8756-8765

Chapter 3

"Artificial Color Tuning of Firefly Luminescence: Theoretical Mutation by Tuning Electrostatic Interactions between Protein and Luciferin"

Naoki Nakatani, Jun-ya Hasegawa, and Hiroshi Nakatsuji

Chem. Phys. Lett. **2009**, *269*,191-194

Chapter 4

"Theoretical Study of the Iron Sulfur Cluster-Free Hydrogenase (Hmd): What is the Active Center of Hmd?"

Naoki Nakatani, Yoshihide Nakao, Hirofumi Sato, and Shigeyoshi Sakaki

Chem. Lett. **2009**, *38*, 958-959

Chapter 5

"Theoretical Study of Dioxygen Binding Process in Iron(III) Catechol Dioxygenase: "Oxygen Activation" vs. "Substrate Activation""

Naoki Nakatani, Yoshihide Nakao, Hirofumi Sato, and Shigeyoshi Sakaki

J. Phys. Chem. B **2009**, *113*, 4826-4836

Chapter 6

"Multi-State CASPT2 Study of Native Iron(III)-Dependent Catechol Dioxygenase and Its Functional Models: Comparison of Electronic Structure and Ligant-to-Metal Charge-Transfer Excitation"

Naoki Nakatani, Yutaka Hitomi, and Shigeyoshi Sakaki

to be submitted