Quantum-based modeling of protein-ligand interaction: The complex of RutA with uracil and molecular oxygen

Igor V. Polyakov, [a,b] Alexander V. Nemukhin, *[,a,b] Tatiana M. Domratcheva, [a] Anna M. Kulakova, [a] Bella L. Grigorenko [a,b]

[a] Department of Chemistry, Lomonosov Moscow State University, Moscow, 119991, Russia

[b] Emanuel Institute of Biochemical Physics, Russian Academy of Sciences, Moscow, 119334, Russia

*e-mail anem@lcc.chem.msu.ru

Keywords: flavin-dependent enzyme RutA, molecular dynamics, oxygenation, protein-ligand interaction, quantum chemistry

Abstract

Modern quantum-based methods are employed to model interaction of the flavin-dependent enzyme RutA with the uracil and oxygen molecules. This complex presents the structure of reactants for the chain of chemical reactions of monooxygenation in the enzyme active site, which is important in drug metabolism. In this case, application of quantum-based approaches is an essential issue, unlike conventional modeling of protein-ligand interaction with force fields using molecular mechanics and classical molecular dynamics methods. We focus on two difficult problems to characterize the structure of reactants in the RutA-FMN-O_2-uracil complex, where FMN stands for the flavin mononucleotide species. First, location of a small O_2 molecule in the triplet spin state in the protein cavities is required. Second, positions of both ligands, O_2 and uracil, must be specified in the active site with a comparable accuracy. We show that the methods of molecular dynamics with the interaction potentials of quantum mechanics/molecular mechanics theory (QM/MM MD) allow us to characterize this complex and, in addition, to surmise possible reaction mechanism of uracil oxygenation by RutA.

Introduction

Modeling the interaction of proteins with small ligands, in particularly, for the goals of drug design is traditionally performed using molecular mechanics or classical molecular dynamics methods with force field parameters. Application of quantum-based simulation tools is also growing,[1,2] but is still far from the routine use. In this work, we describe quantum-based simulations for a molecular system, which can hardly be handled at the classical level. Specifically, we characterize a reactant complex of a flavin-dependent enzyme RutA from *Escherichia coli* with the molecular oxygen and uracil.

The RutA enzyme catalyzes a first step of uracil degradation in hydroxypropionate, ammonia, and carbon dioxide, allowing bacteria to use pyrimidine rings as nitrogen source[3,4] and belongs to a large family of flavoprotein monooxygenases (FMOs) widely spread across

all kingdoms of life.[5] Mammalian FMOs are indispensable as xenobiotic metabolizing enzymes with a prominent role in drug metabolism.[6,7] Growing evidence points to implication of FMOs in aging, several deceases and metabolic pathways.[8] FMOs activate molecular oxygen and insert a single oxygen atom in small organic substrates containing heteroatoms and convert the second oxygen atom in a water molecule.[9]

The chemical formulae of the reactants in the monooxygenation reaction of uracil by RutA are shown in Fig. 1. According to the current knowledge,[10,11] the reduced form of flavin mononucleotide (FMN), which is non-covalently bound to the RutA macromolecule, activates the dioxygen molecule by transferring an electron from FMN. The reactive oxygen is capable to bind chemically to one of the atoms of the isoalloxazine ring of FMN, and this intermediate flavin-based species is involved in the chain of transformations of the uracil molecule. There are unresolved puzzles in this plausible mechanism; one of them is specification of a target atom in the isoalloxazine ring (C4a, N5, C6) for the attack of the reactive oxygen.[10,11] Detailed consideration of the reaction mechanism is far beyond the scope of this work; here, we focus on computational characterization of reactants, namely, on how the ligands, FMN, uracil and O_2, are accommodated inside the protein favoring possible reaction pathways.

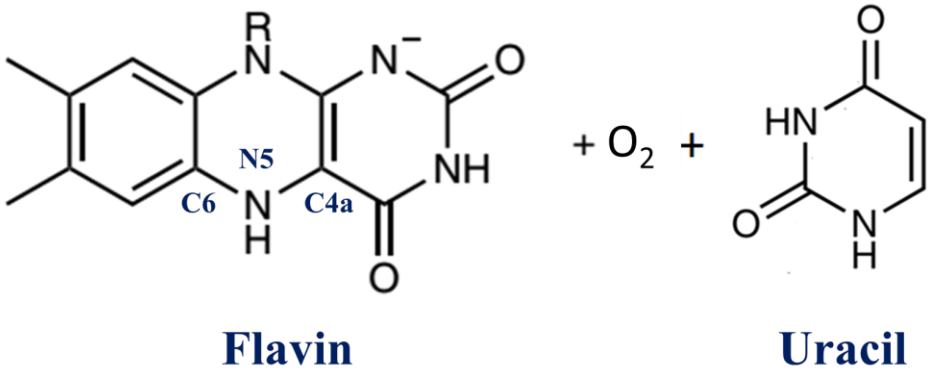

Figure 1. The reactant molecules in the monooxygenation reaction of uracil by the reduced form of flavin mononucleotide.

The work of Matthews et al.[10] reported atomic coordinates of the RutA-FMN-O_2-uracil complex deposited to the Protein Data Bank[12] (the structure PDB ID 6TEG). This complex was obtained as the O_2-pressurized crystal at 15 bars O_2; however, low occupancies of the ligands were observed. From the theoretical side, it should be noted that methods of describing protein-ligand interaction using customary force field parameters may fail in such project, even if some hints are provided by results of crystallography studies. It is not an essential problem to accommodate FMN in the protein cavity, despite location of this ligand close to the solvent accessible area. A small organic molecule, uracil, can also be localized in the active site; however, with more uncertainties, given the disparity in the FMN and uracil sizes. The greatest problem presents the dioxygen molecule because no reliable classical force field parameters are available to describe polarization and a partial charge transfer to this species when it approaches the negative charged FMN and the protein groups. Needless to remind that the localization of dioxygen in protein cavities is a nontrivial experimental task. Thus, quantum-based computer simulations are the indispensable tools to characterize complexes of proteins with small organic molecules and dioxygen.

We apply in this work the advanced modeling method – molecular dynamics simulations with the interaction potentials of quantum mechanics/molecular mechanics (QM/MM MD), which is becoming available recently.[13,14] Classical MD simulations are also utilized at preliminary steps.

Models and methods

To prepare molecular model systems we utilized coordinates of heavy atoms from the structures PDB ID 6TEG and PDB ID 6SGG[10]. The first structure refers to the RutA-FMN-O_2-uracil complex, which we sometimes call below for brevity as the ROU complex. The second structure corresponds to the complex without uracil, RutA-FMN-O_2 (the RO complex). Hydrogen atoms were added assuming the conventional protonation states of the polar residues at neutral pH - Arg, Lys (positively charged); Glu, Asp (negatively charged); the histidine residues were assumed in the neutral state. Water molecules resolved in the crystal structure were kept in the model system, and solvation water box was built using the visual molecular dynamics (VMD) program.[15] The CHARMM36 force field[16] topology and parameters were employed along with the FMN parameters in the reduced form (total charge -3) taken from Ref.[17] and those for the molecular oxygen taken from Ref.[18], while water molecules were treated as TIP3P. Classical MD trajectories were simulated with the NAMD 3.0 software package.[19] The isothermal–isobaric (NPT) ensemble at P = 1 atm and T = 300 K using the Nosé-Hoover Langevin piston pressure control and the Langevin dynamics were employed. Periodic boundary conditions and the particle mesh Ewald algorithm to account for the long-range electrostatic interactions were applied, whereas the non-bonded interaction cut-off parameter was set to 12 Å; integration step was set to 1 fs. A harmonic constraint potential of 1 kcal/Å^2 was applied to the CA atoms of the protein, except for the residues #290-310 of the region that poorly resolved in the crystal structures. In sum, over 500 ns of combined trajectories were produced. The QM/MM MD trajectories were computed with the TeraChem/NAMD via an appropriate interface.[14,20] Calculations were initiated from selected frames of classical MD. The energies and forces in QM parts were computed with the density functional theory (DFT) with the range-separated ωb97X functional[21] and the D3 dispersion correction.[22] The 6-31G** basis set was employed for carbon and hydrogen atoms and 6-31+G** for nitrogen and oxygen atoms. The MM part was treated by the CHARMM36 force field,[16] water molecules were described by the TIP3P model.

Results

Figure 2 illustrates an important result of this work revealed in classical molecular dynamics simulations. According to these simulations either without or with the uracil molecule in the active site, the oxygen molecule can be trapped in several binding pockets in the protein; five of them are schematically shown in Fig. 2a. The retention time of oxygen in these pockets varies from hundreds of picoseconds to tens of nanoseconds. It is important to note that the overall protein structure is stable along trajectories, as shown in Fig. 2b for the protein backbone. The RMSD fluctuations through the graph are mainly due to the region of amino acids #290-310, which is not properly resolved in the crystal structures. The FMN coordination and position is very stable for more than 400 ns trajectory (Fig. 2c), while uracil position fluctuates during the trajectory (Fig. 2d). The hump in the uracil RMSD chart (Fig.

2d) during the ~200-250 ns window is associated with oxygen entering the pocket-5 (Fig. 2a) and consequently pushing off the uracil molecule.

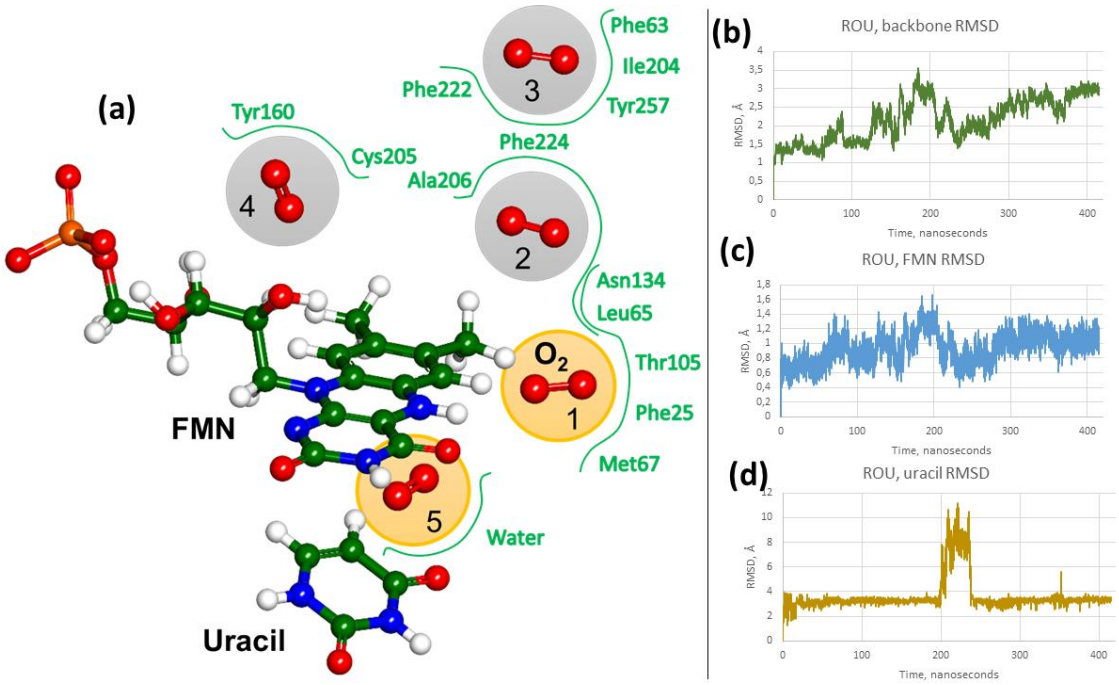

Figure 2. (a) The view on the reactants and the located pockets for the O_2 molecule distinguished by the circled icons. (b) The protein backbone RMSD along trajectories. (c) The RMSD for the heavy atoms of the FMN moiety with the protein backbone aligned along trajectories. (d) The uracil heavy atoms RMSD with the protein backbone aligned along trajectories.

We remind that the starting coordinates for trajectory runs were selected from the crystal structures PDB ID 6SGG and PDB ID 6TEG. The identified pocket-1 roughly corresponds to the structural motifs in the crystal structures. In this site, the oxygen molecule fluctuates in the region between the flavin moiety near atom N5 of flavin (see Fig. 1), polar residues Asn134, Thr105 and a hydrophobic cluster of Leu65, Met67 and Phe25 sidechains (see Fig. 2a). Here, the oxygen and uracil molecule are located at different sides of the plane of the isoalloxazine ring (see Fig. 3 below). In pocket-5, the O_2 molecule is surrounded by water molecules and resides closer to the C4a atom of the flavin molecule. Here, the oxygen and uracil molecule are located at the same side of the plane of the isoalloxazine ring. In other pockets, the oxygen molecule stays farther from the FMN ligand. Pocket-2 is mostly hydrophobic; it is formed by the sidechains of Asn134, Phe224, Ala206, Leu65 and flavin. Pocket-3 consists of the sidechains Phe222, Phe224, Tyr257, Ile204, Phe63; it is close to the protein surface, showing one of the obvious ways connecting the pocket-1 to the bulk. Pocket-4 is formed by the FMN phosphate group, the Tyr160, Cys205 sidechains and nearby main chain of beta-strands. Pockets 1 and 5 are the most perspective sites for the uracil oxygenation reaction in RutA. Therefore, selected frames along classical MD trajectories in the areas

corresponding to pocket-1 and pocket-5 were taken to initiate trajectories with the QM/MM potentials.

In QM/MM MD calculations for the RutA-FMN-O_2-uracil system, with the oxygen molecule in pocket-1, the QM part included the FMN, O_2 and uracil moieties and the Asn134, Thr105 sidechains. Simulations were carried out for triplet spin state of the system due to the ground electronic state of O_2. Correspondingly, the unrestricted DFT approach was used in QM/MM calculations of energies and forces.

Figure 3a shows a general view of the RutA-FMN-O_2-uracil model system. Models of the reactants are presented in right side of Fig. 3. Fig. 3b reproduces the features in the crystal structure PDB ID 6TEG. According to the reported results,[10] the oxygen molecule stands closely to the N5 atom of flavin, the distance between N5 and the nearest oxygen atom is 2.09 Å. The authors of Ref.[10] advocate the reaction mechanism of uracil oxygenation by RutA via formation of a temporary adduct with the O(oxygen)-N5(flavin) chemical bond. Correspondingly, a short distance of 2.09 Å in the reactants resolved in the crystal structure benefits this hypothesis.

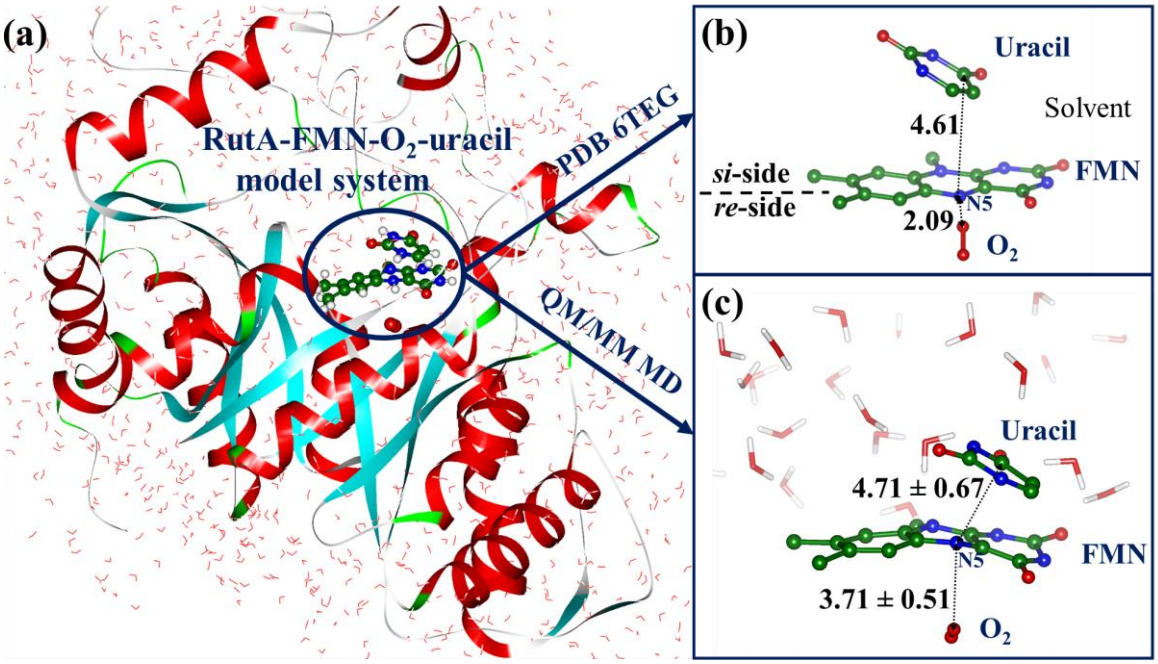

Figure 3. (a) A general view on the model system RutA-FMN-O_2-uracil. The encircled area includes the reactants in the configuration corresponding to the pocket-1 oxygen binding site. The right panels show the reactants in the crystal structure PDB ID 6TEG (b) and in the model system along the QM/MM MD trajectory (c). Molecular models in panels are shown without hydrogen atoms. In this and in all figures, carbon atoms are colored green, oxygen – red, nitrogen – blue, phosphorus – orange, hydrogen - white. The distances are given in Å, showing for the MD results the average values and standard deviation.

However, comparison of Fig. 3b and Fig. 3c shows a strong discrepancy between computational (3.71 ± 0.51 Å) and crystallography (2.09 Å) data for the distance between the oxygen molecule and the N5 atom of flavin. We comment that the reported value 2.09 Å for the crystallography structure PDB ID 6TEG[10] seems to be too short for the hydrogen bond, if

N5 is protonated in the reduced from of FMN, and too long for a covalent bond if one assumes a covalent binding of oxygen to the deprotonated nitrogen in FMN. We remind that location of a small molecule O_2 in protein cavities is a difficult experimental task; therefore, the results of computational approaches should be taken seriously.

As shown in Fig. 4a, the oxygen molecule is localized almost equally close to the N5 and C6 atoms of flavin. Fig. 4b illustrates that the O_2 molecule is stably located in this pocket; this conclusion holds for both complexes, ROU and RO, that is with or without uracil molecule present in the system.

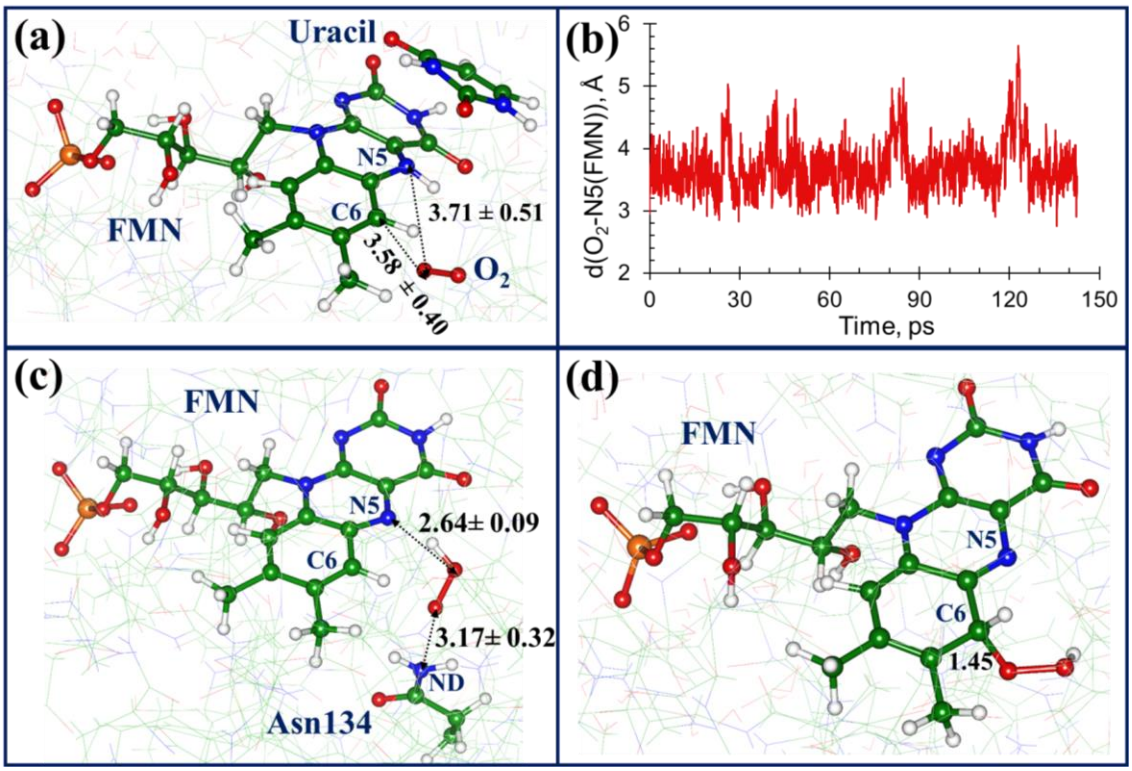

Figure 4. Features of the RutA-FMN-O_2-uracil complex in the conformation around around pocket-1 oxygen binding site. (a) The FMN-uracil-O_2 fragment of the model system in the triplet state. (b) The distance between the center of O_2 and the N5 atom of flavin along the QM/MM MD trajectory (triplet state). (c) The fragment of the model system in the singlet state described by the unrestricted DFT approximation. (d) The fragment of the model system in the singlet state described by the restricted DFT approximation.

At this point, we speculate about subsequent chemical transformations in the system. As mentioned above, when the oxygen molecule comes close enough to the isoalloxazine ring, the electron transfer (coupled with the proton transfer) might occur, resulting in creation of the reactive oxygen species capable to interact with flavin. We model this step by switching the trajectory at some QM/MM MD frame to the singlet spin state of the system. In the case of the unrestricted singlet state DFT approximation, the resulting electronic structure of the system can describe the radical pair FMN and OOH, as shown in Fig. 4c. This structure remains stable along QM/MM MD trajectories, and the Asn134 amino acid residue is very important to stabilize the HOO radical. Interestingly, the average N5(flavin)-O(O_2) distance is very short

(2.64 Å) and exhibits low variation (±0.09 Å), We note that in this singlet radical pair configuration, the N5-O distance is somewhat close (2.64 Å) to the value 2.09 Å, suggested upon the interpretation of experimental density in the PDB ID 6TEG and 6SGG structures.

If the restricted singlet state DFT approximation is applied instead of the unrestricted version, the electronic structure cannot describe radical pairs. Figure 4d illustrates the consequences of such modeling showing that the covalently bound adduct is formed after spontaneous binding oxygen to the C6 atom of flavin. It is not evident how the reactive species shown in Fig. 4c and Fig. 4d react with the uracil molecule given its position on the other side of the isoalloxazine ring; further cumbersome calculations are required to clarify the full reaction pathway. We only note that the main difference between the ROU and RO systems are due to uracil blocking water access to the C6 and N5 atoms of the flavin and to the oxygen atoms of O_2 or subsequent HOO radical/group.

Now we turn to the results of simulations in the area corresponding to the pocket-5 oxygen binding site. In this case, the QM subsystem included FMN, O_2, uracil, the side chains of Trp139, Lys39, Glu143, Asn134 and several water molecules. The results for the QM/MM MD trajectory (Fig. 5) show high variability in the position of O_2 position near the flavin moiety (C4a[FMN]-O1[oxygen]) and the uracil molecule (O1[oxygen]-N3[uracil]). We also note that during the trajectories, oxygen can come close to the Lys69 sidechain - (O2[O_2]-NZ[Lys69]).

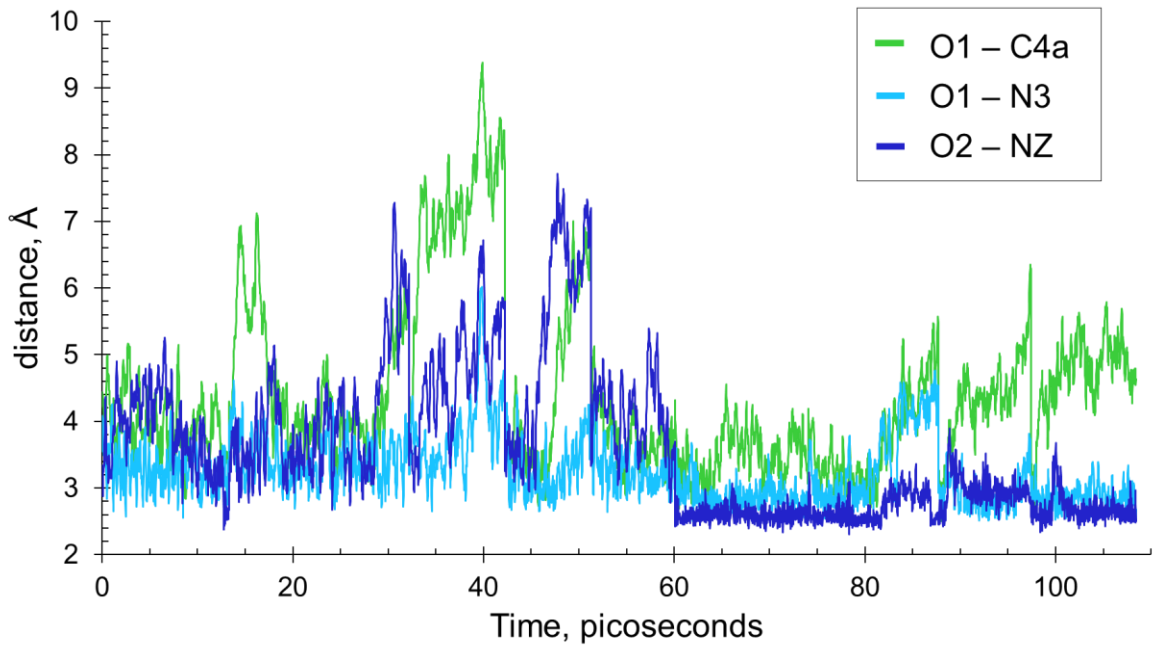

Figure 5. Key distances fluctuations in the pocket-5 during 110 ps QM/MM MD trajectories. Shown are the distances between the oxygen atoms (O1 or O2) to the C4a atom of flavin, the N3 atom of uracil, or the NZ atom of Lys69.

Analyzing these results, we observe significant charge transfer to the oxygen molecule at many steps of the simulation. Such structures are characterized by: (i) close distances of oxygen to the flavin, (ii) close position of the Lys69 sidechain, (iii) proper position of the uracil

that allows for hydrogen bonding network to stabilize the charge transfer. All these issues are observed in the 60-80 ps window (see Fig. 5, Fig. 6a).

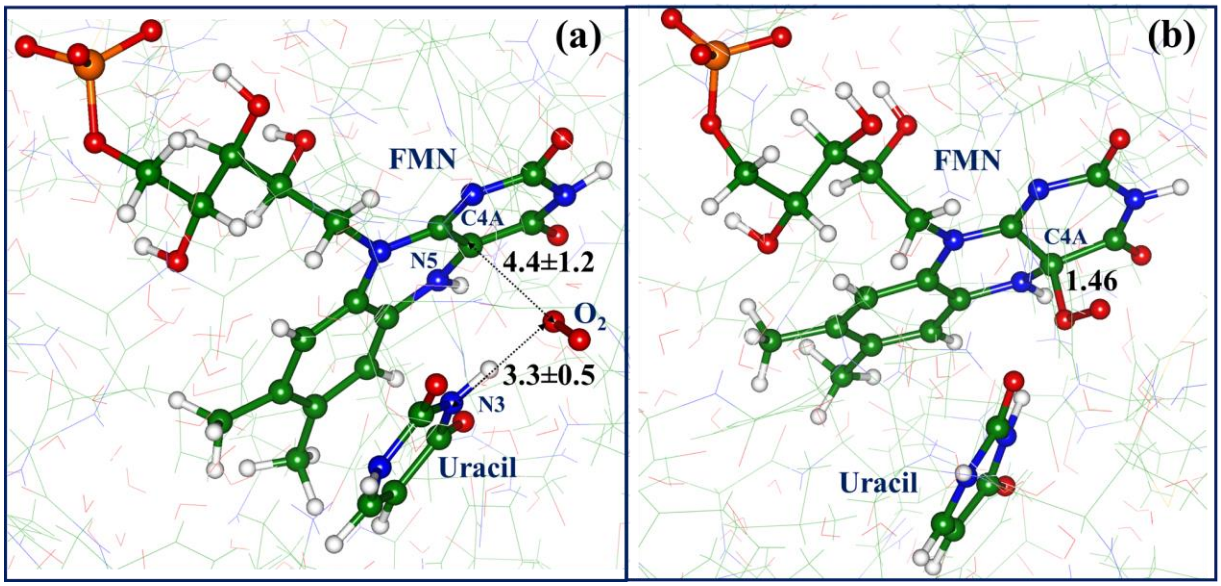

Figure 6. The FMN-O_2-uracil fragment in the model system around pocket-5 oxygen binding site considered in the triplet state (a) and in the singlet state (b).

Here, we also model a possible reaction step toward the intermediate formed by flavin and oxygen. We again switched the trajectory at some frame to the restricted singlet spin state of the system. Figure 6 shows that the covalently bound adduct is formed after binding oxygen to the C4a atom of flavin. Subsequent reactions may lead to the oxidation of uracil.

Conclusions

The results of the simulations described in this work allow us to conclude that the use of the QM/MM MD approaches considerably expands the repertoire of theoretical tools to model complex protein-ligand interaction. The treatment of protein-dioxygen complexes can hardly be handled at the level of molecular mechanics with conventional force fields. As shown in this work, quantum-based approaches may greatly assist in interpretation of crystallography data; in this case, accommodation of the oxygen molecule near the flavin moiety is improved. The QM/MM MD calculations enable us to speculate on the reaction mechanisms when analyzing trajectories near the reactant configurations.

Acknowlegements

We thank the Russian Science Foundation (project 22-13-00012) for the financial support of this work. The study was carried out using the equipment of the shared research facilities of HPC computing resources[23] at M.V. Lomonosov Moscow State University.

References

[1] C. N. Cavasotto, N. S. Adler, M. G. Aucar, *Frontiers in Chemistry* **2018**, *6*, DOI 10.3389/fchem.2018.00188.

[2] O. A. Arodola, M. E. Soliman, *Drug Design, Development and Therapy* **2017**, *Volume 11*, 2551–2564.

[3] K. D. Loh, P. Gyaneshwar, E. Markenscoff Papadimitriou, R. Fong, K.-S. Kim, R. Parales, Z. Zhou, W. Inwood, S. Kustu, *Proceedings of the National Academy of Sciences* **2006**, *103*, 5114–5119.
[4] T. Mukherjee, Y. Zhang, S. Abdelwahed, S. E. Ealick, T. P. Begley, *J Am Chem Soc* **2010**, *132*, 5550–5551.
[5] W. J. H. van Berkel, N. M. Kamerbeek, M. W. Fraaije, *Journal of Biotechnology* **2006**, *124*, 670–689.
[6] J. R. Cashman, J. Zhang, *Annual Review of Pharmacology and Toxicology* **2006**, *46*, 65–100.
[7] S. K. Krueger, D. E. Williams, *Pharmacology & Therapeutics* **2005**, *106*, 357–387.
[8] R. Rossner, M. Kaeberlein, S. F. Leiser, *Journal of Biological Chemistry* **2017**, *292*, 11138–11146.
[9] P. Chaiyen, M. W. Fraaije, A. Mattevi, *Trends in Biochemical Sciences* **2012**, *37*, 373–380.
[10] A. Matthews, R. Saleem-Batcha, J. N. Sanders, F. Stull, K. N. Houk, R. Teufel, *Nature Chemical Biology* **2020**, *16*, 556–563.
[11] S. Adak, T. P. Begley, *Biochemistry* **2017**, *56*, 3708–3709.
[12] H. M. Berman, *Nucleic Acids Research* **2000**, *28*, 235–242.
[13] S. Ahmadi, L. Barrios Herrera, M. Chehelamirani, J. Hostaš, S. Jalife, D. R. Salahub, *International Journal of Quantum Chemistry* **2018**, *118*, e25558.
[14] M. C. R. Melo, R. C. Bernardi, T. Rudack, M. Scheurer, C. Riplinger, J. C. Phillips, J. D. C. Maia, G. B. Rocha, J. v Ribeiro, J. E. Stone, F. Neese, K. Schulten, Z. Luthey-Schulten, *Nature Methods* **2018**, *15*, 351–354.
[15] W. Humphrey, A. Dalke, K. Schulten, *Journal of Molecular Graphics* **1996**, *14*, 33–38.
[16] R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, A. D. MacKerell, *Journal of Chemical Theory and Computation* **2012**, *8*, 3257–3273.
[17] A. Aleksandrov, *Journal of Computational Chemistry* **2019**, *40*, 2834–2842.
[18] S. Wang, K. Hou, H. Heinz, *Journal of Chemical Theory and Computation* **2021**, *17*, 5198–5213.
[19] J. C. Phillips, D. J. Hardy, J. D. C. Maia, J. E. Stone, J. v. Ribeiro, R. C. Bernardi, R. Buch, G. Fiorin, J. Hénin, W. Jiang, R. McGreevy, M. C. R. Melo, B. K. Radak, R. D. Skeel, A. Singharoy, Y. Wang, B. Roux, A. Aksimentiev, Z. Luthey-Schulten, L. v. Kalé, K. Schulten, C. Chipot, E. Tajkhorshid, *The Journal of Chemical Physics* **2020**, *153*, 044130.
[20] S. Seritan, C. Bannwarth, B. S. Fales, E. G. Hohenstein, C. M. Isborn, S. I. L. Kokkila-Schumacher, X. Li, F. Liu, N. Luehr, J. W. Snyder, C. Song, A. V. Titov, I. S. Ufimtsev, L. Wang, T. J. Martínez, *WIREs Computational Molecular Science* **2021**, *11*, DOI 10.1002/wcms.1494.
[21] J.-D. Chai, M. Head-Gordon, *Physical Chemistry Chemical Physics* **2008**, *10*, 6615.
[22] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *The Journal of Chemical Physics* **2010**, *132*, 154104.
[23] V. V. Voevodin, A. S. Antonov, D. A. Nikitenko, P. A. Shvets, S. I. Sobolev, I. Yu. Sidorov, K. S. Stefanov, V. V. Voevodin, S. A. Zhumatiy, *Supercomputing Frontiers and Innovations* **2019**, *6*, 4–11.