

**Reaction Path Analysis of Histone Tail Lysine Residue
Demethylation using only-QM and hybrid QM/MM methods**

by

Bora Karasulu

**A Thesis Submitted to the
Graduate School of Engineering
in Partial Fulfillment of the Requirements for
the Degree of**

**Master of Science
in
Computational Sciences and Engineering**

Koc University

September 2010

Koc University
Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Bora Karasulu

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

Burak Erman, Ph. D. (Advisor)

Özlem Keskin, Ph. D. (Co-Advisor)

Ersin Yurtsever, Ph. D.

Atilla Gürsoy, Ph. D.

Mehmet Sayar, Ph. D.

Date:

ABSTRACT

Chromatins, the basic structural units of the genetic material, consist of DNA and histone proteins. Eukaryotic DNA is wrapped around the histone proteins and they form 'bead-like' structures. Histone proteins control many crucial cell regulatory processes, e.g. gene transcription, gene silencing, DNA replication and repair and etc., via post-translational modifications. Among the post-translational modifications, methylation was recently shown to be reversible by the discovery of Lysine-Specific Demethylase (LSD1) enzyme. As many previous studies have shown the relation of some cancer types and other diseases with the abnormalities in the balance of methylation/demethylation, drug molecule design based on the information gained from reaction mechanism studies becomes very crucial for the fight against these diseases.

In this thesis, a chemically-reliable reaction mechanism is proposed for the demethylation of histone tail lysine residues and the reaction path analysis of this mechanism is carried out. Specifically, demethylation of H3 tail fourth lysine residue, i.e. H3K4, is analyzed. Potential and free energy profiles as well as structural properties are calculated using available Quantum Mechanical (QM), i.e. PM3, B3LYP and MP2, Molecular Mechanical (MM), i.e. UFF, as well as the hybrid QM/MM methods, i.e. ONIOM, which are implemented in *Gaussian09* software package.

As the result of the calculations, it is proved that the proposed chemical mechanism is actually simulating the real-life process and suitable for the demethylation of mono- and dimethylated lysines found on the histone tails. This comment is based on the calculated reaction rates, which are in high agreement with the experimental observations. Besides, some important observations made on the chemical mechanism, which enhances the understanding of how the demethylation process occurs at the molecular level, are explained and discussed in the thesis in fine details.

These results offer an understanding for the details of the reaction mechanism, which will form a fundamental knowledge basis for further studies involving inhibitor molecules. By comparing the standard (i.e. in the absence of any inhibitor molecule) energy profiles and thermodynamic properties with the ones obtained in the presence of the inhibitor candidate, or directly comparing these profiles for different inhibitor candidates, one can have an understanding of the efficiency of the drug candidate for regulating the de/methylation balance. Knowing the key points in the reaction mechanism, one can design novel inhibitor molecules that are inspired by the reaction mechanism.

ÖZET

Genetik materyalin yapıtaşı olan kromatinler DNA ve histon proteinlerinden oluşur. Ökaryotik DNA, histon proteinlerinin etrafına sarılarak ‘boncuk-benzeri’ yapılar oluştururlar. Histon proteinleri translasyon-sonrası modifikasyonlar aracılığıyla gen transkripsiyonu, gen susturulması, DNA eşlenmesi (replikasyonu) ve tamiri, vs. gibi çok önemli hücre düzenleyici süreçleri kontrol edebilirler. Translasyon-sonrası modifikasyonlar arasında, metilasyonun geri döndürülebilir olduğu lisine-özü-demetilaz 1 (Lysine-Specific Demethylase, LSD1) enziminin bulunmasıyla diğer modifikasyonlara nazaran daha geç kanıtlandı. Önceki birçok çalışma metilasyon-demetilasyon dengesindeki anormalliklerin bazı kanser tipleri ve diğer hastalıkların ilişkili olduğunu göstermiştir. Bununla birlikte, reaksiyon mekanizması çalışmalarından elde edilen bilgilere göz önünde bulundurularak ilaç molekülü tasarımı yapmak, bu hastalıklarla mücadele konusunda çok önemli hale gelmiştir.

Bu tezde, histon protein kuyruğunda bulunan lizin amino asitlerinin demetilasyonuna yönelik kimyasal kurallara uygun bir reaksiyon mekanizması önerilmiş ve bu mekanizmanın detaylı analizi yapılmıştır. Özel olarak, H3 kuyruk proteininin dördüncü lizin amino asitinin demetilasyonunu incelemek üzere sistemin potansiyel ve serbest enerji profillerinin yanısıra yapısal özellikleri *Gaussian09* isimli yazılım paketinde bulunan kuantum mekanik (KM) (örneğin, PM3, B3LYP ve MP2), moleküler mekanik (MM) (örneğin UFF) ve melez KM/MM (örneğin ONIOM) yöntemler kullanarak hesaplanmıştır.

Hesapların sonucu olarak, önerilen mekanizmanın gerçek hayatta gerçekleşen tek-metillenmiş ve çift-metillenmiş histon kuyruğu lizin aminoasitlerinin demetilasyonu işlemini açıklayabilen ve ona uygun olan bir mekanizma olduğu, hesaplanan reaksiyon oranlarının (hızlarının) deneysel değerlerle yüksek oranda örtüşmesiyle kanıtlanmıştır. Bununla birlikte, demetilasyon sürecinin moleküler seviyede nasıl işlediğinin anlaşılmasını sağlayacak kimyasal mekanizmaya yönelik bazı önemli gözlemler tezde detaylı bir şekilde açıklanmış ve tartışılmıştır.

Bu sonuçlar, engelleyici (ilaç) moleküllerine yönelik devam niteliğindeki çalışmalar için temel bir bilgi altyapısı kurmuş ve reaksiyon mekanizmasının detaylarının anlaşılmasına bir ışık tutmuştur. Standart (herhangi bir ilaç molekülünün yokluğundaki) enerji profillerini ve termodinamik özelliklerini, bir ilaç molekülünün varlığında elde edilen enerji profilleri ve termodinamik özelliklerle karşılaştırarak yada farklı ilaç moleküllerinin varlığında elde edilen enerji profillerini karşılaştırarak, bir ilacın de/metilasyon dengesinin düzenlemede ne kadar etkili olduğu hakkında bir fikir sahibi olunabilir. Tamamen yeni olan ilaç molekülleri, reaksiyon mekanizmasındaki anahtar noktaların bilinmesi sayesinde tasarlanabilecektir.

ACKNOWLEDGEMENTS

This is the hardest part for me to write throughout this thesis. That is because, it is really not straightforward to thank sufficiently in a single page to all of the people who helped me through writing this thesis.

Firstly, I would like to thank my advisor Prof. Burak Erman for his invaluable guidance and help in all of the issues related to science, thesis as well as private life. I also thank my co-advisor Prof. Özlem Keskin for her help in getting the skill and habit of self-work.

I am extremely grateful to Prof. Ersin Yurtsever, who taught me how to be a theoretician rather than an experimentalist during my undergraduate years, as well as Prof. Atilla Gürsoy and Assistant Prof. Mehmet Sayar for their participation in my thesis committee and for the critical reading of my thesis.

I would like to thank my dear friends, Mustafa ('the Pilot'), Serhat (the '46'), Beytullah and Mümin ('the EU-villagers'), Osman, Besray, Özge, Sefer, Nurcan, Tuğba, Salih and Yasin, who made the two years very enjoyable and helped and guided me in many things without hesitation.

I am grateful to Pelin for loving, supporting and caring me in all of the good and the bad times during those four beautiful years.

The last but not the least, I would like to thank, my big brother Bahadır, my parents Nedim and Naide for their indispensable help, patience, affection, material and spiritual support, and at most for their endless love and care. I am very sure that without them, I would not be in the place I am right now. I love you.

TABLE OF CONTENTS

LIST OF FIGURES	viii
LIST OF TABLES	x
ABBREVIATIONS	xi
PHYSICAL CONSTANTS AND CONVERSION FACTORS	xiii
CHAPTER 1 INTRODUCTION	1
Contribution of this Thesis.....	3
CHAPTER 2 BIOLOGICAL OVERVIEW	4
2.1. Histone Proteins and Post-Translational Modifications.....	5
2.1.1 Information on Histone Proteins.....	5
2.1.1.1 Histone Structure.....	5
2.1.1.2 Linker Histone (H1).....	6
2.1.1.3 Histone Tail.....	7
2.1.1.4 Histone Variants.....	7
2.1.1.5 Epigenetics.....	8
2.1.2 Post-Translational Modifications.....	8
2.1.2.1 Acetylation.....	9
2.1.2.2 Histone Methylation.....	11
2.1.2.3 DNA Methylation.....	15
2.1.2.4 Other PTM Types.....	15
2.2. Lysine Demethylation and LSD1.....	16
2.2.1. Structural and Functional Properties of LSD1.....	16
2.2.2. Jumonji Family Demethylases (JHDMS).....	24
2.3. Relation of De/Methylation with Diseases.....	26
2.4. Chemical Reaction Mechanisms.....	27
2.4.1 Arginine Methylation and Demethylation.....	27
2.4.2 Lysine Methylation.....	29
2.4.3 Lysine Demethylation.....	30
2.4.3.1 Information on Mechanisms Utilized by LSD Family Enzymes.....	30
2.4.3.2 Information on Mechanisms Utilized by Jumonji Family Enzymes.....	35
2.5. Inhibitor (Drug) Molecules Suitable for De/methylation Balance.....	36
2.5.1 Methylation Inhibitors.....	37
2.5.2 Demethylation Inhibitors.....	37
CHAPTER 3 COMPUTATIONAL METHODS AND TOOLS	41
3.1 Aim of this study.....	41
3.2 Literature Review for Theoretical Methods and Computational Tools.....	41
3.2.1 General Information.....	42

3.2.2 Hybrid QM/MM calculations: ONIOM method.....	49
3.2.3 Gaussian Software Package.....	52
3.2.4 AutoDock Software Package.....	56
3.3 Computational procedure used in this study.....	57
3.3.1 Preparation for the Geometry Optimizations.....	58
3.3.2 Geometry Optimizations, Vibrational Analysis and Single Point Energy Calculations using Gaussian09.....	61
3.3.2.1 Optimization Procedure.....	61
3.3.2.2 Frequency Calculations, Stability of the Geometry and High-Level Single Point Energy Calculations.....	63
CHAPTER 4 RESULTS AND DISCUSSION.....	64
4.1 Proposing a Chemical Reaction Mechanism for Demethylation.....	64
4.2 Only QM Simulation System and Calculation Results.....	67
4.2.1 No-Water System.....	67
4.2.2 Single-Water and Two-Water Systems.....	76
4.2.2.1 'Single Water' system.....	76
4.2.2.2 'Two-Water' System.....	78
4.2.3 'O2-Two-Water' and 'MonoMethyl' Systems.....	81
4.2.3.1 'O2-2H2O' System.....	81
4.2.3.1 'MonoMet' System.....	90
4.3 QM/MM Simulation System and Calculation Results.....	95
4.3.1 'OnlyFAD' System.....	95
4.3.2 'Reaction Pocket' System.....	101
CHAPTER 5 CONCLUSION AND FUTURE WORK.....	105
APPENDICES.....	109
BIBLIOGRAPHY.....	121
VITA.....	134

LIST of FIGURES

Figure 2. 1: DNA wrapped around a histone octamer.	5
Figure 2. 2: "Beads on String": The chromatin structure	6
Figure 2. 3: Representation of the binding pockets of LSD1 used to bind and orient the H3 tail accurately.	19
Figure 2. 4: A view from the reaction pocket of LSD1.	20
Figure 2. 5: LSD1 crystal structure in complex with CoREST and H3 tail substrate.	22
Figure 2. 6: Two possible reaction mechanisms for arginine and lysine demethylation	31
Figure 2. 7: Reaction Mechanism for LSD1 demethylation proposed by Shi et al [16].	32
Figure 2. 8: Modified radical mechanism proposed by Silverman et al [92]	33
Figure 2. 9: Formaldehyde recycling scheme proposed by Bannister et al. [99].	35
Figure 2. 10: (i) Molecular structure of tranylcypromine. (ii) and (iii) The adduct structures that tranylcypromine forms with the flavin ring of FAD that is either bound to LSD1 or MAO, respectively.	38
Figure 2. 11: (iii) Propargyl group added histone Lys4 residue. (iv) The adduct complex of Propargyl-added Lys4 and flavin ring of FAD	39
Figure 2. 12: The molecular structure of (-)-(S)-1-Thiocarbamoyl-3-(4-methylphenyl)-5-(4-chlorophenyl)-4,5-dihydro- (1H)-pyrazole	40
Figure 3. 1: Energy surface of hydrogen molecule, H ₂	43
Figure 3. 2: An energy surface example	44
Figure 3. 3: Spin changing phenomenon.	49
Figure 3. 4: QM/MM representation of ethane in ONIOM method.	52
Figure 3. 5: Chemical structures of Flavin Adenine Dinucleotide (FAD), Mono- and Dimethylated lysines	60
Figure 4. 1: Proposed chemical reaction mechanism for conversion of dimethyl lysine into monomethyl lysine via demethylation.	66
Figure 4. 2: The best orientation of dimethyl lysine and flavin obtained with AutoDock using Lamarckian GA.	69
Figure 4. 3: Two proposals for geometry optimization of the transition state structure corresponding to step1 and the converged structure optimized at B3LYP/6-31g* level, +1 charged, TRIPLET	70

Figure 4. 4: Input and optimized geometry for TS structure corresponding to step2 at B3LYP/6-31g* level, +1 charged, TRIPLET.	70
Figure 4. 5: Energy profile for demethylation of dimethyl lysine obtained at PM3 level.	71
Figure 4. 6: Potential and Gibb's Free Energy Profiles of 'R=CH3' and 'Without water' systems obtained at B3LYP/6-31g* level.	73
Figure 4. 7: Singlet and Triplet surface Potential and Gibbs's free energy profiles of 'Without water system'	75
Figure 4. 8: Potential and Gibb's free energy profiles for 'without water' and 'single water' system Obtained at B3LYP/6-31g* level.	78
Figure 4. 9: Potential and Gibb's free energy profiles of singlet and triplet surfaces of 'Two-water' system.	80
Figure 4. 10: A view of the system containing FAD, H3 tail (whose fourth residue is lysine), neighboring enzyme residues, two crystal water molecule and molecular oxygen.	83
Figure 4. 11: B3LYP/6-31g* level triplet surface optimized geometry of step 1 reactants docked by <i>AutoDock</i>	84
Figure 4. 12: Free energy and Potential Energy Profiles of singlet, triplet and quintet surfaces of 'O2-2H2O' system (see text for definition) obtained at B3LYP/6-31g* level.	87
Figure 4. 13: Energy profile for 'O2-2H2O' system obtained at MP2/6-31g* level. Three energy surfaces are present: Singlet, triplet and quintet.	90
Figure 4. 14: Potential and Gibb's free energy profiles of singlet, triplet and quintet surfaces of 'MonoMet' system obtained at B3LYP/6-31g* level.	93
Figure 4. 15: Energy profile for 'MonoMet' obtained at MP2/6-31g* level.	94
Figure 4. 16: Potential and Gibb's Energy profiles for 'onlyFAD' system.	100
Figure 4. 17: A view from step 1 reactant of 'Reaction Pocket' QM/MM system.	102
Figure 4. 18: Potential and Gibb's free energy profiles for 'Reaction Pocket' system obtained at B3LYP/6-31g* level.	103

LIST of TABLES

Table 2. 1: Steady-state kinetic parameters and inhibition of LSD2 and LSD1 obtained at 25 ^o C.....	23
Table 3. 1: Scale factors for some Method/Basis Set for correcting the vibrational frequency and ZPE/Thermal Correction.....	54
Table 4. 1: Optimized structures of the six different models used for finding most probable configuration of the reactant of the first step (Dimethylated lysine and flavin of FAD).....	68
Table 4. 2: The best orientation of aminium radical and FADH [·] obtained with PM3 orientation procedure.....	72
Table 4. 3: Optimized geometries of different orientation trials for ‘Single-water’ system obtained at PM3 level.....	77
Table 4. 4: Two different step 1 reactant structures of ‘Two-water’ system obtained manually and with AutoDock.....	79
Table 4. 5: Optimized geometries of step 1 reactants of ‘O2-2H2O’ system obtained on singlet, triplet and quintet surfaces.....	82
Table 4. 6: Optimized structures for step 1 products of ‘O2-2H2O’ system.....	85
Table 4. 7: Optimized geometries for the two trials for determining the favorable position of single water at the end of step 4 (obtained at singlet surface and B3LYP/6-31g*).....	86
Table 4. 8: Optimized geometries of transition state structures of Step 1 and 2 of ‘O2-2H2O’ with the multiplicity value of 5 (Quintet Surface) and at B3LYP/6-31g*.....	86
Table 4. 9: Optimized geometries of transition state structures of Step 1 and 2 of ‘O2-2H2O’ They are obtained with the multiplicity value of 5 (Quintet Surface) and at B3LYP/6-31g*.....	91
Table 4. 10: Negative frequency values for the optimized TS1 and TS2 geometries of ‘O2-2H2O’ and ‘MonoMet’ systems.....	91
Table 4. 11: Some key structural properties for TS1 and TS2 structures of ‘Without water’, ‘O2-H2O’, ‘MonoMet’ systems.....	92
Table 4. 12: Optimized geometries for Trial 1 and 2 prepared for step 1 products of ‘onlyFAD’ system.....	96
Table 4. 13: Optimized geometry of TS2-like structure for ‘onlyFAD’ system obtained with multiplicity value of 5 and at B3LYP/6-31g*:UFF level.....	98

ABBREVIATIONS

Ac-CoA: Acetyl-CoA
ADMA: Asymmetric N,N'-dimethyl-arginine
AdoHcy: S-adenosyl-L-homocysteine
AdoMet: S-adenosyl-L-methionine
AM1: Austin Model 1
AOF1: Amine-oxidase flavin-containing gene1
AR: Androgen receptor
BLYP: Becke, Lee, Yang and Parr (a DFT method)
CASSCF: Complete Active Space Self-Consistent Field
CCSD: Coupled Cluster Singles and Doubles
CoREST: Co-repressor of REST-dependent genes
DFT: Density Functional Theory
DNA: Deoxyribonucleic acid
DNMT: DNA Methyltransferase
EVB: Empirical Valance Bond theory
FAD: Flavin Adenine Dinucleotide
FEP: Free Energy Perturbation method
G09: Gaussian 09 software package
GA: Traditional Genetic Algorithm
HAT: Histone Acetyltransferase
HDAC=Histone Deacetylase
HF: Hartree-Fock
HFD: Histone Folding Domain
HMT: Histone Methyltransferase
IRC: Intrinsic Reaction Coordinate
JHDM: Jumonji (JmjC-domain-containing) Histone Demethylases
LGA: Lamarckian Genetic Algorithm
LS: Local Search Algorithm
LSD: Lysine-specific demethylase
LSDA: Local Spin-density Approximation
MAO: Mono Amine Oxidase
MD: Molecular Dynamics

MECP: Minimum energy crossing point.
MNDO: Modified Neglect of Diatomic Overlap
NAD: Nicotinamide adenine dinucleotide
ONIOM: Our Own N-layer Integrated molecular Orbital molecular Mechanics
PAO: Poly Amine Oxidase
PcG: Polycomb Group protein
PCPA: trans-2-Phenylcyclopropylamine, also known as tranylcypromine
PDB: Protein Data Bank
PKMT: Protein Lysine-specific Methyltransferases
PM3: Parametrized Model number 3
PRMT: Protein Arginine-specific Methyltransferases
PTM: Post-Translational Modifications
QM/MM: Quantum Mechanics/Molecular Mechanics
REST: Repressor element 1-silencing transcription factor
RMSD: Root Mean Square Deviation
SA: Simulated Annealing
SAH: S-adenosyl-homocysteine
SAM: S-adenosyl-L-methionine
SCF: Self-Consistent Field
SDMA: symmetric N,N'-dimethyl-arginine
TMO: Tryptophan 2-Monooxygenase
TS: Transition state
UFF: Universal Force Field
ZPE: Zero-Point Vibration Energy

PHYSICAL CONSTANTS and CONVERSION FACTORS

Constants:

Planck's constant (h) = $6.62606896 \times 10^{-34}$ J.s

Boltzman constant (k_B) = 1.380654×10^{-23} J.K⁻¹

Avogadro's Number = 6.0221367×10^{23}

Gas Constant (R) = 1.9858775 cal.K⁻¹.mol⁻¹

Conversion Factors:

1 cal = 4.184 J

1 Hartree = 1 atomic unit (a.u.) = 627.46 kcal/mol

1 eV = 23.06037 kcal/mol

CHAPTER 1

INTRODUCTION

Chromatins, the basic structural units of the genetic material, are made of DNA and histone proteins. Eukaryotic DNA is wrapped around the histone proteins and they form a ‘bead-like’ structure. In order to access the genetic information, such as during DNA replication, gene expression, transcription and DNA repair, chromatin is partially unwound [1]. Histone proteins consist of core and tail parts and regulate the gene expression by regulating chromatin structure via post-translational modifications made on their tail parts [2, 3]. These modifications are done at the N-terminus of the specific lysine or arginine residues at specific histone tail and each modification site is related with an important cell regulatory process. For example, H3K4 (Fourth lysine residue on the tail part of core H3 protein), H3K9, H3K27, H3K36, H3K79, and H4K20 are found to be linked to chromatin and transcriptional regulation and DNA damage response [4].

Among the most common post-transcriptional modifications, e.g. acetylation, methylation, ubiquitination, phosphorylation and sumoylation, methylation has been long thought to be irreversible, as the half-life of methylated histone tails are longer than the wild-type (non-methylated) histones [5]. There have been two different views on the mechanism of methylation. According to one, methylation is an irreversible (non-removable) epigenetic mark, which is a heritable change and transferred to the next generations [6-11]. The other side defended that the demethylation is possible [12, 13, 14]. In 1973, Paik and Kim presented the experimental results proving the demethylase activity in rats’ kidney [15]. However, Paik and Kim could not isolate the active enzyme. The relatively late discovery of LSD1 (lysine-specific demethylase), the first demethylase enzyme, in 2004 by Shi et al. [16] clearly showed that methylation modification is, indeed, dynamic (being reversible).

Meanwhile, aberrant methylation/demethylation rates have been shown in the literature to be related with many diseases, especially with some cancer types, heart diseases, diabetes and some neurodegenerative disorders, such as Parkinson’s disease and Alzheimer’s disease [17]. In this frame, designing inhibitor (drug) molecules that are suitable for the de/methylation process is important, as they can be used for regulating the abnormal activities of the de/methylation balance. At this point, there are several inhibitor molecule candidates

available in the literature designed for regulating the abnormal level of de/methylation processes. These inhibitor molecules target either the LSD1 (or flavin ring) or the substrate (target methylation site).

Understanding how the enzymes work is an important question in biology. In order to understand the details of the working chemical mechanism of the demethylation process, a chemically-plausible reaction mechanism is required. Clear understanding of the mechanism makes it available to design novel inhibitor molecules, which are compatible with the mechanism. In this light, novel inhibitor molecule candidates can be provided for the further experimental tests and finally for the clinical use.

In order to assure that the proposed mechanism is appropriate for the demethylation process and working as intended, the reaction mechanism must be modeled using computational methods. Unstable species, (e.g. transition states and/or reaction intermediates) are taking the major role in catalytic activity of the enzyme and have very short life, so that they cannot be detected via experimental screening methods, e.g. X-Ray crystallography, NMR and etc. These unstable species can be modeled via computational methods, making them crucial in modeling and simulating reaction mechanisms.

In this framework, quantum chemical theories are used for a persuasively long time in the literature for modeling enzymatic activity studies. Namely, Density Functional Theory is a widely-used (some example references: [18, 19, 20]) and reliable method that is used to calculate the important thermodynamics properties (e.g. Electronic, free-energy, entropy-profiles) and vibrational properties (e.g. NMR-, IR-spectra) of a given reacting system. Quantum mechanical (QM) methods are used instead of molecular mechanics (MM) methods in modeling reacting systems, as QM methods can imitate the electronic interaction between the atoms and represent the changes during bond breaking and forming that are occurring during the enzymatic reaction.

Relatively recently, the QM/MM concept was introduced in order to study the enzymatic reactions. This concept involves the application of QM and MM calculations simultaneously for different parts of the same reaction system. In this light, QM is applied on the reaction pocket, where the reaction -bond breaking and/or forming- occurs, while the rest of the enzyme residues and other molecules, which are not directly involved in the reaction,

are treated by MM. This provides speed-up in the simulation process and makes it available to take in to account the effect of the enzyme residues on the reaction, as well.

This thesis study consists of five sections and it is organized as follows.

Chapter 2 presents an extensive biological overview. In this chapter, background information is given on histone proteins, post-translational modifications, demethylation process and demethylases (with the focus on LSD1 enzyme). In addition, this chapter includes the reaction mechanisms and inhibitor molecules available in the literature.

Computational methods and tools are discussed in Chapter 3. Namely, overview of the QM and QM/MM theories and optimization techniques are given and computation packages that are used for simulation calculations are explained. In addition, the computational procedure followed for obtaining the results that are given in Chapter 4 are explained in details in this chapter.

Chapter 4 conveys the results of the calculations and the results are discussed in details in this chapter. The thesis is concluded in Chapter 5 and some suitable future work is also given in this chapter.

Contribution of this Thesis

This thesis study presents the fundamental reaction path analysis of the demethylation process of histone tail lysine residues at quantum mechanical level. At first, a chemically reasonable reaction mechanism is proposed and energy profiles and other thermodynamic properties for this proposed reaction mechanism are calculated. The purpose is to determine if the proposed mechanism is probable for the demethylation process, therefore to understand how the demethylation reaction proceeds on histone tails. The effect of the enzyme (i.e. LSD1) residues on the reaction is sought by conducting both only-QM and QM/MM hybrid calculations and comparing the results.

As the result of this study, it is shown that the proposed mechanism is actually working for the demethylation of mono- and dimethyl lysines found on the histone tails based on the reaction rates calculated at MP2-level, which are in high agreement with the experimental results. Besides, calculated potential and Gibb's free energy profiles show that the reaction is exothermic, thus energetically favorable, and spin-changing phenomenon is observed to occur during the reaction.

CHAPTER 2 BIOLOGICAL OVERVIEW

Aim of this chapter is to form a solid biological background that will be indispensable for understanding the work done in the scope of this thesis that will be conveyed in the next chapters. For this purpose, this chapter will start by giving general information on the structural and functional properties of histone proteins and by introducing post-translational modifications (PTMs). In addition, among several PTM types, methylation process is chosen to be explained in details. The focus of the chapter will be brought to the enzymes for methylation and for the reverse of methylation, i.e. methyltransferases and demethylases, and then as the scope of the thesis requires, the chemical mechanisms that are utilized for realizing the methylation and demethylation processes will be explained in sufficient details. Finally, the relationship between the de/methylation balance and important diseases will be discussed and some candidate molecules that can be used for inhibiting the methyltransferases and demethylase activity will be investigated.

Compared to other widely-investigated PTM types, e.g. phosphorylation, acetylation, etc, based on the fact that it was discovered to be reversible very recently [16], methylation is relatively more open and hot topic, so that demethylation has many undiscovered aspects. Methylation has been also shown to be related with the regulation of cell regulatory DNA-based events, such as gene transcription and DNA repair. In that sense, its abnormal levels in the cell, i.e. overexpression or downregulation, causes many crucial diseases, such as tumor and some cancer types, which are also explained in details in this chapter.

In the frame of these important features of LSD1, in this thesis, the focus will be placed on the histone lysine de/methylation balance and LSD1 demethylase. For this purpose, chemical mechanism of the demethylation of histone lysine (i.e. H3K4 or H3K9) catalyzed by LSD1 will be investigated in details. The information gained from this investigation will create the basis knowledge to design novel mechanism-based inhibitor molecules against the diseases explained in this chapter.

2.1. Histone Proteins and Post-Translational Modifications

In this subchapter, structural and functional properties of histone proteins will be given and some general and important points on post-translational modification types will be explained.

2.1.1 Information on Histone Proteins

2.1.1.1 Histone Structure

In almost all eukaryotes, DNA is wrapped around histones to form nucleosomes that secure the genome and makes it available to be accessed during active processes, such as transcription, replication and DNA repair. Each nucleosome consists of 145-147 bp of DNA wrapped around a protein octamer that is made up of two of the canonical (core) histone proteins: H2A, H2B, H3 and H4 [21]. Each of the core histones has the Histone Fold Domain (HFD) and tail parts in common. The HFD is made up of $\alpha 1$, $\alpha 2$ and $\alpha 3$ helices and L1 and L2 loops and holds the antiparallel complementary pairs -H3 and H4; H2A and H2B- together in order to form the dimeric structure. These dimeric structures come together to form tetramers and finally the histone octamer (See Figure 2.1). Octameric histones are found in all eukaryotes except dinoflagellates –marine protozoa with two flagella-, whose bulk histones were determined to be disappeared a while ago [22]. Nucleosome cores are bound to each other via compact linear DNA in order to form high-order structure and they are stabilized by the linker histone protein, i.e. H1. [21].

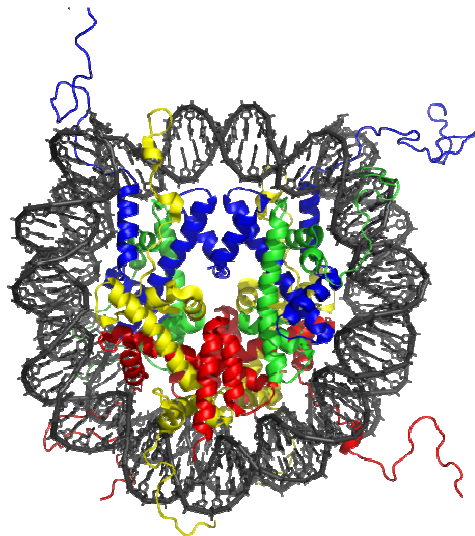


Figure 2. 1: DNA wrapped around a histone octamer. Color code for core and tail histones is given as: Yellow: H2A, Red: H2B, Blue: H3, Green: H4 [23].

Histone Fold Motif (HFD) is highly conserved in a variety of organisms, e.g. archaeal bacteria, insects, birds and amphibians, because of its DNA-binding properties. Histone sequences are less conserved except the HFD, i.e. core and tail parts, in order to form the different characteristics of DNA superhelix, which fulfills the higher-order structure requirements and provide the interaction with assembly-, transcription- and remodeling factors [22].

In electron microscopy studies, the chromatin structure has been shown to be so-called “beads on string”, which is shown in Figure 2.2. In this model, linker DNAs are called “string” and the histone octamers are called “beads”.

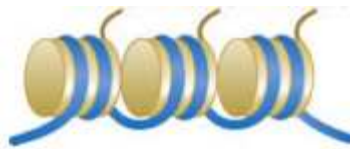


Figure 2. 2: “Beads on String”: The chromatin structure [24].

Core histone proteins are highly conserved and important in modulating chromatin structure and DNA accessibility for replication, repair and transcription [3].

2.1.1.2 Linker Histone (H1)

Linker histone in human has a conserved globular domain and relatively longer C-terminal tail and shorter N-terminal tail [25]. Linker histones are crucial for cell growth and division [26]. As H1 links the nucleosomes to each other, reduction in H1 amount results to a significant change in mammal chromatin structure. The compactness of chromatin depletes, thus the nucleosome-repeat length decreases. Reduction of H1 also causes depletion of H3K28 acetylation, H3K28 trimethylation, and also DNA methylation [26].

Linker histones have been found to be controlling the epigenetics via PTMs. For example, acetylation of the N-terminus of H1 linker histone is the most abundant PTM. This modification is correlated with the replication rate in human and mouse cells. Another example is phosphorylation. Phosphorylation of the linker histones is an indicator of the mitosis and mainly involved in the cell-cycle control [27]. To be more specific, some phosphorylation sites (e.g. Ser18, Ser173 and Ser189) of H1 linker histone of human lymphoblastic T-cells are related with interphase and mitosis stages [28]. Therefore, phosphorylation of Linker H1 is related with growth rates of the organism.

In contrast, abundance of lysine methylation on linker H1 may be attributed to the rapid proliferation requirements of the corresponding cell. In relation to this fact, aberrant linker histone methylation is reported in HeLa cells [25].

2.1.1.3 Histone Tail

Tail part of the core histones make up the 28% of the mass of the core histone proteins [22]. N-terminal tails of the core histones can be up to ca. forty-residues long and they extend through the nucleosomal packing. Therefore, for rearrangement of the chromatin structure via PTMs, the nucleosome packing needs not to be unwound completely, as the PTMs are done on the tails outside the nucleosomal packing. The characteristics of the tails are constantly modified by the PTMs. C-terminus of the tails are also modified by PTMs, but they are not as frequent as the N-terminal are.

Until now, H3 takes the lead among the core histones by having the largest number of modifications [25]. It is followed by H4, H2B and H2A, respectively.

2.1.1.4 Histone Variants

Histone proteins are altered by simultaneous modifications done on more than one residue. Eukaryotic histone variants are widely found in almost all eukaryotes, e.g. centromeric histone variant (CenH3), core histone variants (H3.2, H3.3, H2A.Z and H2A.X), and linker histone variants (H1.2, H1.4 and etc). Among core histones, only H3 and H2A are variant, while H2B and H4 are generally invariant.

The histone variants are essential, as they are involved in DNA repair, meiotic recombination, chromosome segregation, transcription initiation, sex chromosome condensation, sperm chromatin packaging, and developmental regulation via differentiating the structure and thus altering the stability of the nucleosomes [22]. For example, CenH3 is involved in formation of kinetochore, which is the protein complex found on the centromere, on which the microtubules of the mitotic spindle are attached during the mitosis [29].

Similarly, H3.3 is the replacement of canonical H3 (also called H3.1), whose only four amino acids are different from the H3.1, i.e. A31S and other substitutions at the positions: 87, 89 and 90. H3.3 has been shown to be mediating the epigenetic memory of an activated gene state via its fourth lysine residue [22].

Some histone variants (e.g. protamines, H3t, TH2B, TSH2B, etc.) are involved in packaging of animal sperm DNA. Their specific role has not been determined yet, however their role is thought to be erasing the paternal epigenetic states and/or facilitating the mobilization of the sperm by forming less stable nucleosomes [22].

In addition, H1.2 and H1.4 are the variants of linker histone H1. They are very similar to each other, having a sequence identity of 93.6%, but they show great diversity from the wild-type H1. Linker histone variant H1.2 is shown to be involved in translocation to the cytoplasm, thus leading to permeability of the mitochondrial membrane and release of the material created by apoptosis. Some studies showed that H1.2 is infected with a virus causes increased mobility and this may play an important role in cancer treatment [25].

On the other hand, linker histones and their variants are also modified by PTMs [25]. For example, Lys46 in H1.2, H1.3 and H1.4 in human HeLa cells, Lys116 of H1.1 and Lys46 of H1.2 and H1.3 in mouse tissue HeLa cells have been shown to be ubiquitinated. In contrast, Lys88 on H1.5 and Lys90 on H1.1, H1.2, H1.3 and H1.4 of human MCF97 cells were shown to be formylated [30].

2.1.1.5 Epigenetics

Epigenetics is the heritable changes in gene expression that is not encoded in DNA sequence but transferred to the next generations. Epigenetics is defined and controlled by RNA silencing, DNA methylation and post-translational modifications (PTMs) that are applied on the core and linker histone tails [25]. These modifications govern a multitude of genomic functions, particularly gene expression, and are considered to constitute an epigenetic code, which governs many nucleus-regulatory functions, e.g. gene expression, X-chromosome inactivation, DNA repair and replication.

These functions are realized either directly by changing nucleosome interactions within chromatin or indirectly by recruiting effector proteins, such as CHROMO, TUDOR, MBT, PHD domains and etc, each of which recognizes some specific covalent modifications (i.e. PTMs) [31].

2.1.2 Post-Translational Modifications

Gene activation is a dynamic process, so that it can rapidly respond to the physiological changes. In contrast, DNA needs to be compact; therefore it is wrapped around histone

proteins, forming the nucleosome structure, which is transcriptionally repressive. For providing transcriptional responses to the physiological changes, the nucleosomal repression must be overcome. At this point, post-translational modifications, which are covalent changes by nature, take the stage and change the chromatin structure without spoiling or unwinding its compactness. In that sense, information diversity that leads to human complexity is stemming from the post-translational mechanisms [3].

A post-translational modification (PTM) is the covalent addition of a chemical group to the N-terminal or C-terminal of the target residue. The target residue is found either on the tail of the target core histone, linker histone or the histone variant. There are many discovered types of PTMs. Acetylation, phosphorylation, methylation, ubiquitination, formylation, sumoylation and ribosylation are the major PTM types, which have been widely investigated so far. The target residues are of many types. Modification types and the corresponding residue types, on which the PTMs are applied, are as follows: Acetylation on lysine, phosphorylation on serine and threonine, methylation on lysine and arginine, sumoylation on lysine and ubiquitination on lysine residues [25].

In the literature, it was proven that the activity level of one PTM type is affecting the activity level of another PTM type. For example, H3 tail Lys9-Ser10 and Lys27-Ser28 are adjacent methylation-phosphorylation sites and they are involved together in gene transcription. They regulate the activity of each other in a tandem manner [32]. Likewise; acetylation is shown to be regulating the methylation/demethylation level. They are not given here in details, but interested reader may refer to the references [33] and [34]. In that sense, it will not be wrong to say that different PTM types work cooperatively by making cross-talks in order to regulate cellular processes.

There are many PTM types and much accumulated information on these PTM types. Explaining all of them in full-details requires great effort and will be inconclusive. Therefore, for the sake of keeping the scope of the thesis narrow, only two of PTM types, i.e. acetylation, and methylation, will be given in sufficient details as required by the scope of the thesis.

2.1.2.1 Acetylation

Acetylation is the transfer of the acetyl group from acetyl-CoA (Ac-CoA) to the ϵ -amino nitrogen atom of histone lysine residue in the presence of acetyltransferases (HATs) as the catalyst. After the first acetyltransferase discovery in 1996 by Allis et al [35], acetylation

process has been extensively studied and many acetyltransferase families have been identified. There are four major families of HATs, i.e. GNAT, MYST, p300/CBP and Rtt109, based on the sequence conservation [17, 36]. Among these families, GNAT family is conserved from yeast to human, while p300/CBP and Rtt109 is specific for metazoan and fungal, respectively.

HATs have some interesting properties to point out. HAT enzyme families have great sequence diversities, such that p300/CBP family HATs show no sequence homology to other HAT family enzymes, except the acetyl-CoA-binding motif that is common in all HAT families [36]. Another interesting property is the auto-acetylation of p300/CBP, and Rtt109. These HATs, themselves, are acetylated in order to increase their acetylation activity [36].

p300/CBP, GNAT and MYST family enzymes have been shown to follow a similar chemical mechanism, which is an ordered bi-bi kinetic mechanism involving a ternary complex intermediate. The mechanism, which is given in Appendix C3, differs only in terms of the substrate specificity and it is analyzed in details in the given references [17, 36]. The chemical mechanism of the other family, i.e. Rtt109, has not been determined yet.

Human GNAT and p300/CBP HATs were shown to acetylate the non-histone substrates, e.g. p53 tumor suppressor and DNA binding proteins, although MYS and Rtt109 have not been shown yet to acetylate non-histone substrates [36].

The reverse of acetylation is called deacetylation and it involves the removal of the acetyl group from the N-terminus of the acetylated lysine residue and catalyzed by histone deacetylases (HDACs). The first deacetylase was discovered in 1996 by Taunton et al. [37]. HDACs are divided into four classes based on the phylogenetic analyses: Class I, II, IV and Sirtuin (Class III) [17]. Class I HDACs contain human HDAC1-3 and HDAC8, class II HDACs contain HDAC4-7 and HDAC9-10, and class IV HDACs include human HDAC11. Class II HDACs (sirtuins) are conserved from bacteria to human and some yeast homologs exist [17].

Class I/II/IV HDACs are considered to follow similar reaction mechanism, which is given in Appendix C4, as they contain homologous active-site motifs. This mechanism recruits an active-site metal, either Zn(II) or Fe(II), which is coordinated by the active-site enzyme residues, i.e. histidines, asparagine, and tyrosine, and the reaction proceeds through a tetrahedral intermediate. In contrast, class III HDACs (sirtuins) are considered to utilize a distinct, nicotinamide adenine dinucleotide (NAD)-dependent mechanism, which is given in

Appendix C5. This mechanism proceeds through a ternary complex, which is formed by NAD and conserved active site enzyme residues [17].

2.1.2.2 Histone Methylation

Methylation is simply the covalent addition of a methyl (CH₃) group to N-terminus of target core or linker histone tail. Its reverse process is called demethylation, which is simply the removal of a methyl group from the methylated site. Methylation can be done on arginine or lysine residues on the tails of core and linker histones and also on the tails of histone variants. The target lysine or arginine residue, on which the methylation occurs, is called the methylation site. A methylation site can be mono, di or tri-methylated, each of which indicates how many methyl group is attached to the target nitrogen atom of the methylation site. Lysine methylation sites can have all of three methylation degrees, while arginines can only be mono, and dimethylated. Each degree of methylation corresponds to a different function [38-40]. Methylation sites are abbreviated based on the methylation degree and the target residue. For example, H3K4me2 stands for the fourth dimethylated lysine residue that is found on the tail of core H3 protein.

Besides, the relatively high stability of lysine methylation compared to other modifications, e.g. acetylation, methylation, is considered to be stemming from the fact that the N-CH₃ bond is thermodynamically highly stable [41]. In contrast to acetylation, histone methylation does not change the charge of the histone's target residue, rather it changes the basicity, and hydrophobicity of the histone tail and also its affinity to some specific proteins, e.g. transcription factors [42].

Lysine methylation is involved in gene silencing (especially H3K10, H3K28, and H4K21) and active gene transcription (H3K5, H3K37, and H3K80) [25]. The number of methyl transferases increases significantly from lower eukaryotes to higher eukaryotes (by increasing complexity). Based on the global chromatin organization human has more methylation sites compared to yeast. All eukaryotes own the methylation marks that are related to transcriptional activation, i.e. H3K4, H3K36, and H3K79, while the budding yeast does not have any methylation mark that is characteristic for condensed chromatin and transcriptional repression, e.g. H3K9, H3K27, and H4K20. In higher eukaryotes, each degree of methylation is controlled via different enzymes, while in lower eukaryotes all of them are controlled via only single enzyme [39].

Contrary to acetylation and phosphorylation, methylation was shown to be reversible recently. Before the discovery of the first demethylase [16], there was a controversy on if methylation is a reversible process or, in other words, if methylation sites are permanent marks. One side thought that the histone methylation is not reversible, as the turnover rates of methyl lysine are low; therefore methylation marks can be demethylated passively by natural histone turnover or by DNA replication [11, 43], and active demethylation is not possible. This thought was compatible with some types of transcriptional silencing. That means, a stable methyl group that has long-term-conserved state was useful as an epigenetic mark for the inheritance of the silenced state [38]. As Shi et al. explain in their study [16], before the discovery of LSD1, the natural methylation turnover was thought to be occurring in two different ways by the scientists who considered the methylation as static process. The first one involved the clipping of few N-terminal residues of the methylated histone tail by the cell, while the other one suggested replacement of the methylated histone with a variant histone, as histone variants differ from the original histones only by a few number of amino acids [38].

On the other hand, some scientist defended that methylation is reversible by introducing some studies (before the discovery of LSD1), in which the active turnover of histone methylation was detected at a very low but detectable level [44-46]. The first demethylase activity was observed in the 1960s by Kim et al. in their study, in which the authors purified an enzyme from rat kidney that shows the demethylation activity of free mono- and dimethyl lysines [47]. A few years later, the same scientists could partially (i.e. in very low ratio) purify this enzyme, but could not allocate the demethylase activity to a specific protein [15, 48]. However, until the discovery of the LSD1 in 2004 [16], the presence of a histone demethylase could not be proven.

Compared to other modification types, e.g. phosphorylation, acetylation and etc; histone methylation sites has a lower global turnover rate, i.e. methylated histones exist longer compared to the bulk histones. Therefore, they are considered to be involved in the establishment and propagation of different gene expression patterns in the same genome, thus they are known as *epigenetic marks* [39]. This means heritable changes in gene expression are carried to the next generations through germline via methylation sites. The discovery of the demethylases [16] has lead to doubts about accepting histone methylation sites as epigenetic marks. To be unambiguous, it was thought that if methylation sites are reversible, then they

cannot be epigenetic marks, which are accepted as non-reversible, stable and consistent marks. Trojer and Reinberg [39] support that this is not the case by alleging the following points. The presence of histone demethylases does not increase the turnover rates of methylation sites, making them remain stable. In addition, although DNA methylation has been shown to be reversible, its epigenetic activity has been still observed and proven. Another point is that a biological function may not be constricted to the message carried in one type of methylation mark. Therefore, although being reversible, methylation sites can still be epigenetic marks.

An important indicator that a methylation site is an epigenetic mark is that it has to be established and maintained throughout the cell cycle. In this frame, monomethylation of H4K20, for example, fulfills this requirement. In details, H4K20me1, which is monomethylated by PR-SET7 enzyme, is not removed by lysine demethylases in the interphase cells during the early embryonic development; and deficiency of which results in lethality [39]. Mitosis is proposed to be the final step for passing the epigenetic information from parental chromatin to the next-generation chromatin before the chromosome segregation.

On the other hand, EZH2 and PR-SET7 are important epigenetic regulators, which methylate the H3K27 and K4K20 methylation sites [39]. H3K27 is methylated by EZH2, which is a member of Polycomb (PcG) protein family and is critical for maintaining transcriptional repression of homeotic (HoX) genes. Polycomb (PcG) family proteins have been shown to be regulating the proper embryonic development and involved in maintaining of cellular identity. In that sense, EZH2 shows its importance in establishing cellular identity, which is an example of epigenetic memory, based on its H3K27 methylase activity [39]. EZH2 can also interact directly with DNA methyltransferase; therefore it has an indirect effect on epigenetic phenomena via DNA methylation.

H3K9 methylation is associated with heterochromatin formation and euchromatic gene repression. H3K9 is methylated by Suv39H and then bound to chromatin domain protein HP1 [16]. On the other hand, H3K4 methylation is related with active transcription, as is methylation of the arginine residues of H3 and H4. H3-K4 specific methylase was shown to cooperate with RNA polymerase II [16].

2.1.2.2.1 Methyl Transferases

Histone methyltransferases (HMTs) are responsible for methylation of target residues on lysine or arginine residues. Lysine- and arginine-specific methyltransferases are abbreviated

as PKMTs and PRMTs, respectively. The first PKMT was discovered in 2000 by Jenuwein et al. [49] and they are dependent on S-adenosyl methionine (also known as SAM and AdoMet). There are two families of lysine methyl transferases, SET domain containing proteins and non-SET domain containing DOT1/KMT4 analog enzymes. There are also the arginine methyltransferases.

SET domain enzymes have over 60 discovered enzymes that are found in human. Based on the active site residue pattern, this family enzymes can carry out mono-, di- or trimethylation of the target methylation site. [17]. The SET domain contains 12 β -strands bifurcated by an inserted variable region, i.e. iSET, that is involved in substrate recognition. The iSET region is flanked by nSET and cSET regions. The cSET region plays an important role in substrate binding and catalysis [31].

PKMTs show narrow substrate specificities, i.e. they target single histone methylation site. For example, SET8 and DIM-5 demethylate only H4K20 and H3K9, respectively. In contrast, SET7/9 shows broader substrate specificity, e.g. H3K4, Lys 372 of p53 and Lys189 of TAF10. Compared to other PKMTs (e.g. SET8 and DIM5), a shorter binding motif is observed in SET7/9, which is conserved when SET7/9 is bound to H3K4, p53 and TAF [31].

Some examples of SET domain-containing PKMTs are DIM-5 in *Neurospora*, human SET7/9, human SET8 (also known as PR-SET7), and vSET in *Paramecium bursaria* chlorella virus [31]. In addition, EZH2 and mixed lineage leukaemia (MLL) are the SET-domain PKMTs that methylate H3K4 methylation site and they were shown to be related with cancer [50].

Dot1/KMT4 family enzymes do not contain a SET domain and can methylate Lys79 on the H3 core [17]. They can only methylate nucleosomal histones, not the free histones and they have AdoMet binding motifs similar to those of histone arginine methyl transferases family.

Dot/KMT4 enzymes were shown to have a cross-talk relationship with ubiquitination of core histone H2B [2]. When Lys120 residue on human H2B is ubiquitinated, the H3K79 methylation rate was shown to increase. However, how the ubiquitinated H2B stimulates H3K79 methylation has not been determined yet. There are some proposals for that mechanism which are given in [51].

Although Dot1/KMT4 methyltransferases do not contain a SET domain, they utilize similar methylation mechanism to SET-domain containing enzymes, which will be explained in subchapter related to reaction mechanisms.

2.1.2.3 DNA Methylation

DNA methylation is a reversible epigenetic modification that regulates genomic stability and cellular plasticity. Normal DNA methylation levels are required for healthy development in mammals.

DNA methylation occurs at the 5-carbon position at the pyrimidine ring of the cytosine base, which precedes a guanine base in a CpG dinucleotide sequence. This modification is simply covalent addition of a methyl (CH₃) group to the carbon atom -described above- by the catalysis of DNA methyltransferases (DNMTs), in which S-adenosylmethionine is used as the methyl source. The reverse of methylation is the removal of methyl group and it is called demethylation. Demethylation was shown to be directly involved in DNA repair [52].

PTMs on the core and linker histones influence the level of DNA methylation. For example, acetylation of H3K10 is in reverse-correlation with DNA methylation amount [53]. Similarly, presence of H1 inhibits the DNA methylation [25].

DNA methylation has been shown to be related with many crucial diseases, which will not be given in this thesis. Nevertheless, the reader, who is interested in the DNA methylation-disease relationship, is suggested to refer to the reviews [25, 54].

2.1.2.4 Other PTM Types

Besides acetylation, phosphorylation and methylation, there are other types of post-translational modifications that attract many investigators' research attention, e.g. sumoylation, ubiquitination, formylation and etc. These modification types apply on core histones, linker histones and histone tails and they are also involved in some regulatory processes and regulated by some enzymes. Besides, the abnormal levels of these modification types are also related with diseases. No detailed information on them will be given for narrowing down the scope of this thesis; nevertheless some very useful references are given for the reader, who is interested in these modifications. These references are [17, 36, 55, 56] as general reviews and related to diseases, [57-61] for phosphorylation, [58, 59, 60] for ubiquitination, [59] for sumoylation and [62] for formylation.

In addition, a very informative table is presented by Peterson and Craig [63] that relates most of the modification types and the corresponding sites with the possible function and enzymes involved in catalysis of this modification. Moreover, the review written by Chi, Allis and Wang [64] is very instructive on the relationship of PTMs with human cancer types.

2.2. Lysine Demethylation and LSD1

Before the discovery of the first histone specific demethylase [16], the de/methylation (methylation-demethylation) balance was shown to be present, but any type of specific demethylase could not be observed via experimental methods, e.g. crystallography. In the short period following the discovery of LSD1, an indispensable number of novel histone specific demethylases have been discovered. These demethylases are grouped into two families based on their structural and functional properties, i.e. chemical mechanism, coenzyme use and reaction products [40]. These families are the LSD family, which also contains the LSD1, and the JHDM (JmjC-domain-containing) family.

On one hand, LSD family enzymes can demethylate mono- and dimethyl lysines using an amine oxidation mechanism, which is dependent on FAD and molecular oxygen and produces formaldehyde and water. On the other hand, JHDM enzymes can demethylate trimethylated lysines, in addition to the mono- and dimethyl lysines, using a radical attack mechanism, which requires α -ketoglutarate, Fe(II) and molecular oxygen to proceed and produces formaldehyde, succinate and carbon dioxide.

Detailed information will be given on the structure and chemical mechanisms corresponding to these enzyme families in the following subsections.

2.2.1. Structural and Functional Properties of LSD1

LSD1 (also known as AOF2, BHC110, KIAA0601, p110b, and npao) is the first discovered histone demethylase and the first member of the lysine-specific demethylase (LSD) family, which have only two members discovered until now. The other member is LSD2 that was identified by Karytinis et al. in 2009 [2]. LSD1 is a FAD-dependent amine oxidase and it was discovered by Shi et al. in 2004 [16]. It can demethylate mono- and dimethylated H3 tail Lys4 and Lys9 methylation sites and therefore has an indispensable role in gene repression, gene activation and active gene transcription [65, 66]. It shows no strong kinetic preference

among between the mono- and dimethyl lysines [67]. The crystal structure of LSD1 was determined by two separate groups [66, 68] and structural details will be given below.

LSD1 has a highly asymmetric, closely packed core structure that is accompanied by long helical tower domain. The LSD1 molecule is ca. 60 Å wide and ca. 140 Å high and it is composed of six structural domains [66]: TOWER, SWIRM (Moirra), Oxidase (also called AOL in [68]), Swi3p, Rsc8p and N-terminal region. Amongst them, only C-terminal Oxidase domain has the reaction pocket and thus is related to the demethylation reaction. The other domains affect the demethylation process via allostery. The reaction pocket has sufficient space to accommodate several residues of the histone tail to be modified, implying that the selectivity of the LSD1 to the H3K4 and H3K9 is stemming from chemical mechanism of the reaction, not from the steric hindrance [66]. This is in agreement with what Shi et al. [16] had proposed.

N-terminal SWIRM domain consists of ca. 100 residues and forms a helical histone fold. SWIRM domain is also conserved in other chromatin remodeling and modifying complexes [68]. SWIRM domain is bound to the C-terminal of Oxidase domain. This interface is held together by extensive van der Waals and hydrogen bonding network, and it is highly hydrophobic. SWIRM domain was shown [69] to be involved in DNA binding, since it is conserved in other proteins realizing chromatin remodeling. However, according to Stavropoulos et al. [66], contrary to other histone-modifying enzymes, SWIRM domain is not directly involved in DNA binding. Nevertheless, SWIRM domain is required for the stability of LSD1 and thus for catalysis. This was shown in the study [66] as the mutation of the conserved residues of the SWIRM domain, which are thought by the authors to be related with engaging the histone tail substrate, leads to decrease in enzymatic activity.

The TOWER domain consists of two long anti-parallel helices, which are connected to each other via hydrophobic interactions. The domain is directly connected to the catalytic centre of the enzyme. This domain is thought to be acting as a lever and regulates the catalytic activity allosterically, i.e. without direct intervention [66]. This means, when other proteins (e.g. CoREST, which will be explained later in this section) connect to the TOWER domain, the size of the reaction pocket is regulated. As the result, enzymatic activity is regulated.

Oxidase domain has an amine oxidase fold that is conserved in other flavoenzymes, e.g. Mono Amine Oxidase (MAO) and Poly Amine Oxidases (PAO). In that sense, LSD1 is a

nuclear homolog of amine oxidase (AO) type enzymes [68]. Oxidase domain can covalently bind to the Flavin Adenin Dinucleotide (FAD) molecule, which is required for demethylation process and the substrate, i.e. methylated lysine. The oxidase domain contains the active site for demethylation process. When compared to other MAOs, the active site of LSD1 is more spacious. It consists of ca. 50 highly conserved residues (conserved among different species). The active site can accommodate the residues of the histone tail flanking to the target lysine residue in order to adjust correctly the position of the lysine according to the flavin ring of the FAD molecule [66].

Most of the residues belonging to the catalytic cavity are conserved in close homologs, e.g. LSD2 [68]. The C-terminal 2/3 domain of LSD1 displays a high-level sequence homology to FAD-dependent amine oxidase, e.g. MAO, PAO. Therefore, this sequence homology makes LSD1 available to catalyze the oxidation of amines including monoamine, polyamine or N-methylated protein substrates (e.g. histone) [16]. In that sense, LSD1 can be alleged to be a FAD-binding amine oxidase that adapted two additional domains (i.e. TOWER and SWIRM) in order to act on a specific substrate, i.e. histone tail lysine residue.

LSD1 was also shown by Shi et al. as a transcriptional repressor and it is also involved in restricting neuron-specific gene transcription in non-neuronal HeLa cells. [16]. In addition, dimethylated p53-K370 is necessary for efficient binding to the transcriptional coactivator p53-binding protein-1. LSD1 can control the tumor suppressor activity of p53 via demethylating Lys370 of p53 [70]. At this point, it has not been explained how LSD1 can recognize a non-histone methylation site, which differs from histone lysine methylation target in terms of surrounding residues.

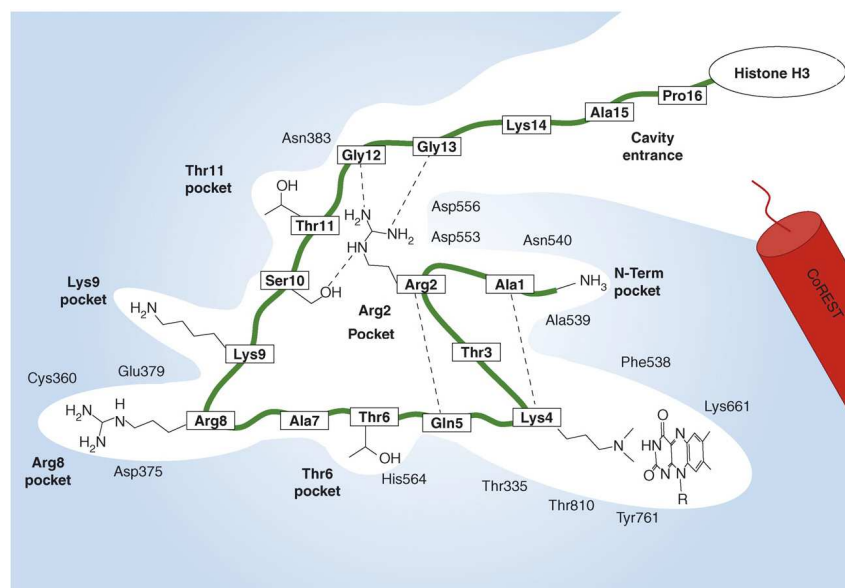


Figure 2. 3: Representation of the binding pockets of LSD1 used to bind and orient the H3 tail accurately. Using the Arg2, Thr6, Arg8, Lys9, Thr11 and N-term pockets, Lys4 residue is held in precise orientation with respect to flavin ring of FAD in the reaction pocket [40].

According to Stavropoulos et al. [66], the oxidase domain contains four pockets for binding to the lysine and the flanking residues of the histone tail. These pockets are shown in Figure 2.3 in a nice graphical representation prepared by Forneris et al. [40]. By the means of these pockets, the methylation state and the position of the lysine (4th or the 9th) can be detected. The first pocket is the catalytic chamber (which is surrounding the isoalloxazine ring of FAD) and it consists of Val317, Gly330, Ala331, Met332, Val333, Phe538, Leu659, Asn660, Lys661, Trp695, Ser749, Ser760 and Tyr761. It has a highly negative electrostatic potential [68]. Among these conserved residues, Lys661 is of significant importance. Lys661 is hydrogen-bonded to a crystal water molecule, which is hydrogen-bonded to the N5 of the flavin ring (isoalloxazine ring). This lysine residue is also conserved in other oxidases, e.g. PAO and MAO-B [71]. This lysine molecule also stabilizes the flavin ring via Pi-Plus (Π^+) interaction. When Lys661 is mutated, the demethylation activity of LSD1 is completely removed [68, 71].

While the catalytic chamber holds the substrate lysine, the other three pockets bind to the flanking residues. The importance of these three pockets was shown by Stavropoulos et al. via mutation experiments, in which the mutation of some chosen residues resulted with significant decrease in enzymatic activity [66].

Although the reaction pocket is spacious enough to accommodate trimethyl lysine, LSD1 cannot demethylate the trimethyl lysine. In contrast, the trimethyl lysine acts as a competitive inhibitor. Therefore, LSD1 cannot discriminate the methylation state of the lysine. This discrimination is made chemically via the reaction mechanism [66], details of which will be given in the following subchapters related to reaction mechanisms. A view of the reaction pocket of LSD1 is given in Figure 2.4. It was prepared using *Accelrys DS Visualizer*.

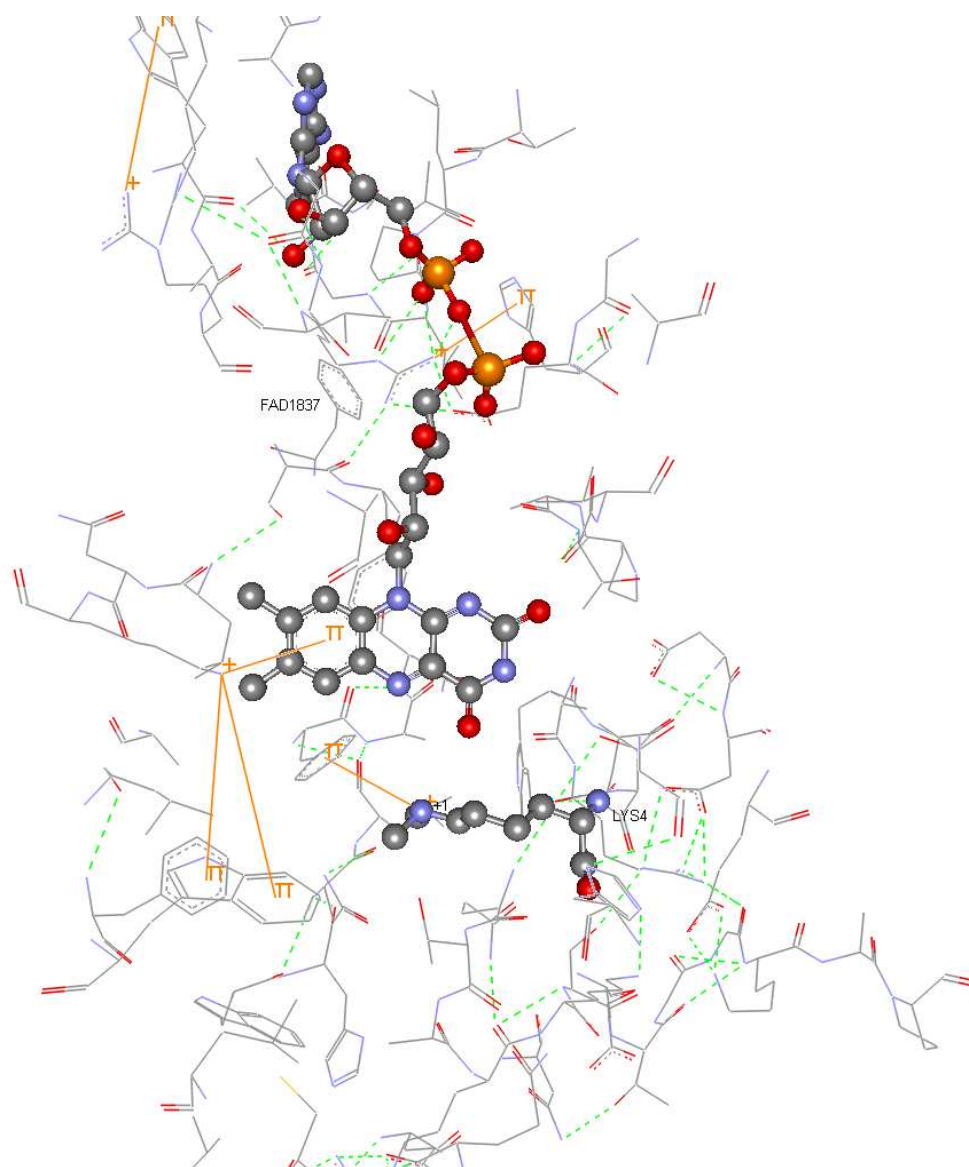


Figure 2. 4: A view from the reaction pocket of LSD1. Derived from PDB ID: 2V1D. H3 tail fourth methionine is mutated back to Lysine. FAD also is present. FAD and Lysine are given in ball-and-stick model, while neighboring LSD1 residues are given in stick representation. Hydrogen bond and Pi Interactions are shown in green dashed lines and orange solid lines, respectively. The lysine is dimethylated and its N-terminus is +1 charged.

LSD1 is a component of transcriptional co-repressor complex (BHC) that is involved in silencing neuronal-specific genes [71, 72]. LSD1 can interact mainly with four different proteins, which are other components of BHC complex and regulate the demethylase activity of LSD1. CoREST is the co-repressor of REST-responsive (RE1-silencing transcription factor) genes and makes LSD1 available to demethylate nucleosomes. BHC80 inhibits the activity of LSD1, while androgen receptor switches LSD1 specificity between H3K4 and H3K9 [39, 65, 71].

Besides histone demethylation, BHC complex also shows histone deacetylase (HDAC) activity via its HDAC1/2 components. BRAF35 is another component of BHC, but it has no direct effect on the demethylase activity of LSD1. However, it can potentiate the deacetylase activity of HDAC on nucleosomal substrates [71].

CoREST consists of three domains, i.e. ELM2, SANT1 and SANT2. SANT2 domain is involved in the interaction of CoREST with DNA (having a dissociation constant, K_D , of 84 μM), while the linker region that is found between SANT1 and SANT2 domains binds to LSD1 TOWER domain via hydrophobic and some ionic interactions [72]. In that sense CoREST acts as a bridge across DNA and LSD1 during nucleosomal histone demethylation. In fact, it was shown by Shi et al. that LSD1 can demethylate the H3-K4 in peptides of bulk histone, however only LSD1-CoREST complex is able to demethylate the H3-K4 within nucleosomes. The reason for this fact is given by Shi et al. as the formation of LSD1-CoREST complex protects LSD1 from proteasomal degradation [65]. CoREST presence also increases H3 deacetylation rate [71].

LSD1-CoREST complex has three main parts: the base that holds the SWIRM and Oxidase domains of LSD1, the stalk consisting of TOWER domain of LSD1 and the CoREST linker region, and the head that is formed by the SANT2 domain of CoREST. The linker region creates three major interfaces between LSD1 and CoREST. The first interface consists of CoREST $L\alpha 1$, LSD1 $S\alpha 1$ and $S\alpha 2$, while the second interface contains LSD1 $I\alpha 1$ and $I\alpha 2$, and CoREST $L\alpha 2$. The third interface is formed of LSD1 $I\alpha 2$ and CoREST SANT2 [72]. In Figure 2.5, graphical representation of the LSD1-CoREST complex is shown.

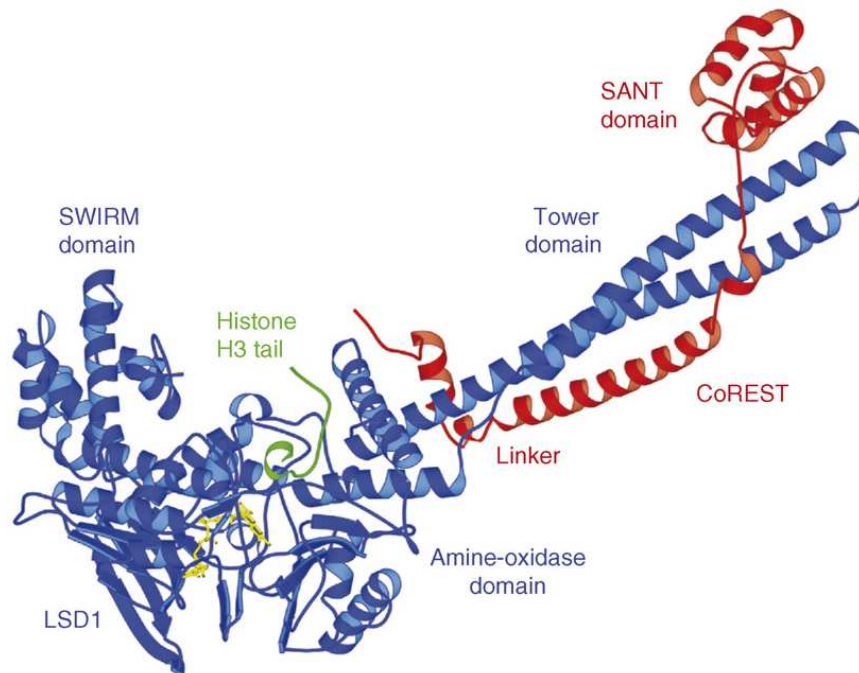


Figure 2. 5: LSD1 crystal structure in complex with CoREST and H3 tail substrate. PDB ID: 2V1D was used for creating the representation. Adapted from [40].

BHC80 is another component of BHC complex and contains PHD-domain and acts as LSD1 inhibitor. It was discovered by Hakimi et al. [73]. As shown in the experiments done by Shi et al. increasing amounts of BHC80 decreased the demethylase activity of only LSD1 and LSD1-CoREST complex [65]. However, the mechanism of how BHC80 inhibits LSD1 activity is unclear.

On the other hand, histone deacetylases (HDACs) are the histone specific deacetylases and increase LSD1 activity in nucleosomal histones. Hyper-acetylated nucleosomes were shown to be less susceptible to CoREST/LSD1-mediated demethylation. HDACs create hypo-acetylated nucleosomes that are better recognized by LSD1-CoREST complex [65]. Therefore, HDACs are thought to be interacting functionally with LSD1 to create a repressive chromatin environment.

LSD1 requires the first 21 N-terminal amino acids of the histone tail in order to bind effectively. This provides LSD1 a more precise substrate binding ability and makes it available to sense epigenetic message encoded by the histone tail. Moreover, presence of other epigenetic messages (e.g. hyperacetylation and phosphorylation) decreases demethylation activity of LSD1. This may infer that LSD1 acts on the histone tails on which only methylation marks are present, and also it may infer that enzymes of other PTM types

need to finish their work on the histone tail before LSD1 can start its task [40]. In that sense, LSD1 works as the cleaner of the last-standing epigenetic mark on the histone tail.

On the other hand, LSD2 is the second and –for the time being- the last member of LSD enzyme family [2]. In mammals, LSD2 enzyme is encoded by the amine-oxidase flavin-containing gene1 (AOF1) that is homologous to the LSD gene. LSD1 and LSD2 show 33% sequence identity in amine-oxidase and SWIRM domains. LSD2 is considered to be utilizing the same chemical mechanism as LSD1 is, since peroxide and formaldehyde production is detected as the result of the demethylation process by Karytinis et al. Similar to LSD1, LSD2 requires and is non-covalently bound to a FAD molecule. In addition, LSD2 shows the maximal enzymatic activity at pH 8.5 and it requires at least 21 amino acids from the N-terminus of H3 tail in order to be able to show its demethylase activity, which is the same case for LSD1 [2]. In contrast, LSD2 shows more or less the same turnover rates working with a peptide with 30 amino acids of H3. The steady-state experimental reaction rates are given in Table 2.1. From this table, it is seen that LSD2 has a higher demethylase activity compared to LSD1 and both LSD1 and LSD2 have a slightly higher turnover rate for dimethylated lysine compared to monomethyl lysine.

From Table 2.1, it is also exerted that mutation of Lys4 to methionine results in highest affinity value for peptide binding to the enzyme. This implies that peptides, which are homologous to H3 tail but whose target lysine site is mutated can be used as inhibitor, as well. The peptide binds to enzyme well (with a high affinity value) and shows no demethylation activity showing inhibitory effect.

Modification type		LSD2			LSD1		
N ^a	Name	K _m (μM)	K _{cat} (min ⁻¹)	K _i (μM)	K _m (μM)	K _{cat} (min ⁻¹)	K _i (μM)
21	Monomethyl Lys4	9.2 ± 0.9	0.28 ± 0.01	-	3.4 ± 0.2	3.40 ± 0.10	-
21	Dimethyl Lys4	11.3 ± 1.3	2.00 ± 0.60	-	4.2 ± 0.5	8.10 ± 0.20	-
21	Trimethyl Lys4	No activity	No activity	58.0 ± 6.6	No activity	No activity	19.5 ± 3.2
16	Monomethyl Lys4	No activity	No activity	-	No activity	No activity	-
30	Monomethyl Lys4	5.1 ± 0.5	0.41 ± 0.01	-	3.4 ± 0.5	2.90 ± 0.10	-
21	Lys4Met	No activity	No activity	0.15 ± 0.1	No activity	No activity	0.05 ± 0.02

Table 2. 1: Steady-state kinetic parameters and inhibition of LSD2 and LSD1 obtained at 25^oC. Values are taken from [2]. N^a is the number of amino acids in the peptide (used instead of H3 tail in the experiments).

Similar to LSD1, LSD2 can sense the PTMs that are applied on other residues on the H3 tail, on which they are working. For example, being acetylated of the other residues (rather than H3K4) impairs the demethylase activity of LSD2 [2]. Moreover, hyper-acetylation of Lys9, 14 and 18 lead to inactivity of LSD2 as demethylase. Similarly, phosphorylation of Thr3 and Ser10 and, separately, methylation of Arg2 and Arg8 cause the loss of demethylase activity of LSD2, which is very similar to the effects observed with LSD1 (Corresponding experimental values can be found in [2]).

On contrary to LSD1, LSD2 is strictly specific for H3K4 and it lacks the TOWER domain. Therefore, it cannot create a stable complex with CoREST. However, contrary to LSD1, LSD2 contains a CW-type zinc finger domain in order to mediate the interaction with DNA and proteins. Having different domains, LSD2 is expected to be involved the chromatin-remodeling complexes and transcription programs that are different from LSD1. For example, LSD2, contrary to LSD1, is shown to be not able to demethylate the Lys370 residue on p53 tumor repressor, as well [2].

2.2.2. Jumonji Family Demethylases (JHDMS)

The first member of the large JHDM (Jumonji Histone Demethylases) family, whose members own JmjC-domain in common, was discovered in 2006 by Tsukada et al. [74]. The JmjC domain is conserved from yeast to human, and there are, for the time being, approximately 30 JHDMS identified in human [17]. Although the first JHDM was discovered more recently compared to the first LSD family enzyme, the number of discovered JHDMS is very high compared to the number of discovered LSD1 enzymes. JHDM family enzymes can demethylate trimethyl lysines in addition to mono- and dimethylated lysines by the means of the chemical reaction mechanism they follow, which will be explained in the subsection related to reaction mechanisms. In this frame, JHDMS and LSD family enzymes may be used in cooperation to remove completely a methylation mark, so that making these epigenetic marks completely reversible.

JMJD2A/KDM4A, the first discovered JHDM, can demethylate mainly trimethylated, and also -to small extent- dimethylated Lys9 and Lys36 sites on H3 tail, i.e. H3K9me_{2/3} and H3K36me_{2/3} [17]. These two methylation sites share no significant sequence homology. This dual site specificity is performed by the methyl ammonium binding pocket that is flanking to the Fe(II) and 2-OG (2-oxoglutarate) binding pockets. This pocket contains an array of

oxygen atoms that make direct contact with the ζ -methyl groups of the methylated lysine substrate via hydrogen bonding [36]. This is analogous to key-lock model, where the lysine and the JMJD2A are the key and the lock, respectively. When the lysine is not methylated or mono-methylated, the key is not complete, thus it does not fit into the lock.

JMJD2A can also demethylate the H3K4 and H4K20 methylation sites by the means of its double tudor domains, i.e. HTD-1 and HTD-2. The target methylation site is held in a cage of three aromatic residues. Two of these residues are coming from the tudor-2 domain, while the last one is from tudor-1 domain, which defines the substrate specificity in cooperation with other residues from tudor-1 domain [75]. Further structural details of JMJD2A, which stay outside the scope of this thesis, can be found in the references [36, 76-78].

Jumonji (JHDM) family enzymes have been alleged to have a turnover rate in the range of 0.01 min^{-1} , which 300-fold lower compared to LSD1 (LSD1 has a reaction rate of 3.4 min^{-1}). Similarly, other Fe(II)/2-oxoglutarate-dependent dioxygenase superfamily enzymes have two to four-fold reaction rates (k_{cat} values) compared to JHDM family enzymes. This unexpected low rate of JmjC enzymes points to the need of auxiliary proteins in order to stimulate JmjC activities [3, 17]. Therefore, this issue requires further investigation as future work [17].

On the other hand, Yjr119Cp in yeast, JARID1C in human and Lid (little imaginal discs) in *D. melanogaster* can demethylate the H3K4me3 methylation mark. Over-expression of Yjr119Cp was shown to be resulting in DNA replication defects and loss of telomeric silencing [24]. Likewise, JARID1b was shown to be demethylating di- and trimethylated H3K4. Therefore, JARID1b may be used complementary to LSD1 in complete removal of H3K4 methylation mark in human. Such cooperation was shown for JMJD2c and LSD1 during androgen receptor-dependent gene expression [79].

Similar to LSD family enzymes, Jumonji family enzymes have been shown to have an indispensable role in important cell regulatory processes, e.g. transcription, chromatin regulation, cell proliferation, cardiac and brain development. Specific examples are conveyed in details in [80]. Therefore, defects in JHDM activity results in crucial diseases.

Information on which demethylase is related with demethylation of which methylation site is given as a table in Appendix D1.

2.3. Relation of De/Methylation with Diseases

Some studies have proven that the abnormal levels of any PTM types cause important diseases. For example, abnormal methylation/demethylation results in cancer [81, 82], heart diseases [83], rheumatoid arthritis [84], diabetes [85, 86], neurodegenerative disorders, e.g. Parkinson's disease and Alzheimer's disease [85, 87] and etc. In this subsection, within the scope of this thesis, only diseases related with methylation and demethylation will be given. The relationships of other PTM types with diseases can be followed from the references provided in the previous subsections.

LSD1 may play a crucial role in tumorigenesis, as it is involved in several growth-promoting pathways and LSD1 levels were shown to be upregulated in high-risk tumors [16, 40]. Loss of H3K4 methylation and enhancement of H3K9 methylation were shown to be linked to several tumor types.

Repression of REST target genes is realized by LSD1-CoREST-HDAC core, through lysine deacetylation on H3 and H4 and demethylation of H3K4. Any abnormalities in the balance of these modifications result in the loss of the REST-mediated repression, thus ectopic reactivation of the neuronal genes [40]. Similarly, *hematopoiesis*, the process of formation of blood cell components, is an important developmental process and requires LSD1-CoREST-HDAC complex as transcriptional corepressor. The regulatory role of LSD1 in *hematopoiesis* is similar to its role in REST-mediated neuronal gene silencing [40]. In that sense, any inconvenience in LSD1 activity result in the disease related to blood cell anomalies.

In addition, LSD1 is able to activate gene transcription. Demethylation of H3K9, dependent on LSD1 presence, is involved in the activation of androgen receptor (AR) target genes. Similarly, LSD1 regulates the level of growth hormone expression during the pituitary development [40]. LSD1 is also considered as a new target for prostate cancer treatment.

Likewise, CARM1/PRMT4, an arginine methyltransferase, is overexpressed in breast tumors and hormone dependent prostate tumors, whereas the overexpression of EZH2 is linked to breast and prostate cancer [42]. The human hDOT1L lysine methyltransferase was shown to be related with cancer and leukemias. Knockdown of hDOT1L expression resulted in inhibition of cell proliferation.

On the other hand, trihorax group protein, i.e. MLL, which is a methyltransferase for H3K4 methylation site, was found to be related with the chromosomal translocation in both

acute lymphoid and myeloid leukemia [16]. Likewise, SMYD3, a H3K4 histone methylase, is proven to be up-regulated in colorectal and hepatocarcinoma cells. In addition, when SMYD3 is overproduced, cell proliferation is increased based on the over-methylation of histone, thus SMYD3 is an oncogene candidate [16].

Moreover, mutation of JARID1C, a JHDM, was shown to be directly leading to X-linked mental retardation. Similarly, JARID1b and other demethylases are misexpressed in tumors. This implies that the demethylases may serve as predictive tumor markers [24].

In Appendix D2 and D3, two tables are given showing the relationship of some arginine and lysine methyltransferases with some known cancer types. In the tables also the references to the related work is given. These tables provide very useful knowledge for the reader who is interested in the methylation related diseases.

2.4. Chemical Reaction Mechanisms

As it was explained in details in the previous subsection, abnormal de/methylation levels lead to vital diseases. Therefore, methylation-demethylation regulation is crucial in preventing the related diseases. For regulating these PTM levels, knowing how these PTMs are working at the molecular level (i.e. chemical mechanism) in details makes it available to design novel inhibitor (drug) and activator molecules that affect the related PTM level.

In this subsection, chemical reaction mechanisms of different types of histone methylation and demethylation will be given. Nevertheless, within the scope of the thesis, only the reaction mechanism of LSD1-catalyzed histone tail lysine residue demethylation will be investigated in fine details. The reaction mechanisms of other PTM types, e.g. phosphorylation, acetylation, formylation, ubiquitination and etc, will not be given in this thesis for narrowing down the coverage of the thesis. However, they are all available in the references placed in the Post-translational Modifications subsection.

2.4.1 Arginine Methylation and Demethylation

Arginine residues of histone proteins are modified via methylation and citrullination (deimination) by the means of protein arginine methyltransferases and arginine deiminases, respectively. Methylation is related to transcriptional regulation, translation and DNA repair [88]. Human prostate carcinoma and coronary heart diseases were shown to be related with over-expression of arginine methyltransferases [89, 90].

Protein arginine methyltransferases (PRMTs) catalyze the methylation of arginine residues on the histone proteins, which require AdoMet as the methyl source. There are two types of PRMTs. Both of the enzyme types can do the monomethylation of the arginine. Type I enzymes, e.g. PRMT1, PRMT3, CARM1/PRMT4, PRMT6 and PRMT8, can form the asymmetric N,N'-dimethyl-arginine (ADMA), while type II enzymes, e.g. PRMT5, 7, and 9, can form symmetric N,N'-dimethyl-arginine (SDMA) [42].

In the active site of PRMT enzymes, a glutamate and an arginine residues are conserved, which interact with carboxylate and two ribose hydroxyls of AdoMet in order to create the reaction channel similar to the histone methyltransferases. Other two conserved glutamate residues make hydrogen bonds to guanidinium side chain of arginine in order to position the lone pair electrons on N-terminus of the arginine towards the methylsulfonium of the AdoMet. This facilitates the attack of the lone pair electrons to the methylsulfonium of AdoMet [17]. The arginine methyl transfer follows an S_N2 mechanism similar to the lysine methyltransferases (see Appendix C1 for chemical mechanism).

The subsequent methylation of demethylated and monomethyl arginines is possible, as is in the case of lysine methylation. This is carried out via the release of AdoHcy and binding of the new AdoMet, while the methylated arginine is released after the second methylation step [17]. The sequential methylation without the release of methylated arginine substrate is considered to be achieved via the dimer structure of two conserved catalytic core regions, each of which contains an AdoMet binding site motif [17].

PRMTs are involved not only in histone arginine residues, but also many other proteins that are related with transduction and cell proliferation. For example, histone acetyltransferase CBP is methylated by methyltransferase CARM1 and through this, CARM1 regulates the co-activating effects. In addition, PRMTs have also been shown to affect RNA maturation and DNA repair. For example, DNA-damage-control protein, MRE 11, is regulated by PRMTs [42].

On the other hand, arginine demethylation is catalyzed by the arginine demethylases, the first of which, JMJD6, was discovered very recently in 2007 by Chang et al. [91]. Chang et al. showed that JMJD6 is a homolog of JHDM family enzymes, which is specific for demethylation of H3R2me₂ and H4R3me₂, but not for H3R17me₂ or H3R26me₂. Being a homolog of the JHDM enzymes, JMJD6 is considered to be following the same reaction

pathway as JHDM enzymes for the demethylation of arginines (See Appendix C2) and its activity was shown by Chang et al. to be dependent on the presence of Fe(II) and 2-oxoglutarate. In the active site of JMJD6, three residues, which are also conserved in JHDMS, are involved in binding to Fe(II). The demethylase activity is also dependent on these residues.

The JMJD6 was also shown to have low arginine demethylation rates, as well as to be related with significant oxidation of two lysine residues, which allege that JMJD6 is not specific for arginine demethylation [17]. For understanding further details on the arginine demethylation, further demethylases should be characterized and investigated. In this frame, investigation of the arginine demethylation in terms of reaction mechanism at molecular level is a hot research topic.

2.4.2 Lysine Methylation

There are two different families of histone methyl transferases (HMTs), as explained in the previous subsection related to lysine methylation. These families use similar chemical mechanisms (reaction pathway) for methylating the target lysine, which is given in Appendix C1.

For SET-containing family enzymes, active site contains a conserved aspartate or glutamate residue that forms hydrogen bonds to ribose hydroxyls of the AdoMet, while it also contains a conserved arginine or lysine residue that form a salt bridge with the carboxylate of the AdoMet. These interactions orient the AdoMet and substrate, i.e. lysine. The C-H...O hydrogen bonds, which are formed between the methyl group and the side chain hydroxyl of Tyr335 residue and carboxyls of Gly264 and His293 residues of SET domain enzyme, are also important in positioning and activation of AdoMet. The most favorable position for methylation is when the sulfur of AdoMet, the carbon of the methyl to be translated and the ϵ -amine of lysine are linear. By the means of this geometry, via S_N2 nucleophilic attack of N-terminus of the lysine to form the product, S-adenosyl-L-homocysteine (AdoHcy) [17]. Methyl transfer (i.e. methylation of lysine) only occurs if ϵ -amine of lysine is deprotonated. Some mechanisms have been proposed for deprotonation, which are deprotonation by an active-site residue, such as Tyr245 or Tyr 335, by an ordered water molecule, by bulk solvent or by decreasing pKa of ϵ -amine through closing it up to the cationic methylsulfonium group of AdoMet [17].

In order to make multiple methylations successively, ϵ -amine must be further deprotonated by rotating the carbon-nitrogen bond and the lone pair electrons, which are formed as the result of deprotonation, are aligned with the methyl-sulfur bond.

Although Dot1/KMT4 HMTs do not possess any SET domain, they occupy similar chemical mechanism as the SET-domain enzymes do [17]. The interaction between Thr139, Gln168, Glu186, Asp161 and Asp222 of DOT1/hKMT4 enzyme found in human and the AdoMet align the methylsulfonium group of AdoMet, creating a channel in which substrate lysine binds. As the same for SET-domain enzymes, S_N2 methylation mechanism occurs by the means of this channel. Similar to SET-domain enzymes, the AdoMet and lysine substrate bind to different sites of the enzyme, so that successive methylation of lysine can be done without the release of lysine residue. In addition, the deprotonation of lysine residue is also required for methylation catalyzed by DOT1 enzymes to start, as it is the same case for the methylation catalyzed by SET-domain enzymes. There are some hypotheses for how deprotonation is done, which are given in details in [17]. It is not known for sure which one of these hypotheses is valid for deprotonation of lysine for DOT1 enzymes, but as anticipated it is different than the deprotonation mechanism for SET-domain enzymes [17].

2.4.3 Lysine Demethylation

2.4.3.1 Information on Mechanisms Utilized by LSD Family Enzymes

As it was already explained in the previous section, before the discovery of LSD1 in 2004, scientists did not agree on if the methylation was a reversible process and if the methylation could be removed by an active demethylation process. However, some scientist argued and showed some cases in which active demethylation was observed. In this frame, even though the presence of a histone demethylase could not be proven, some chemical mechanisms for histone demethylation were proposed. The first proposal was created by Bannister et al. in 2002 [38]. Bannister et al. proposed two different mechanisms for demethylation of arginine and lysine residues. The overall reaction mechanisms (pathways) are given in Figure 2.6.

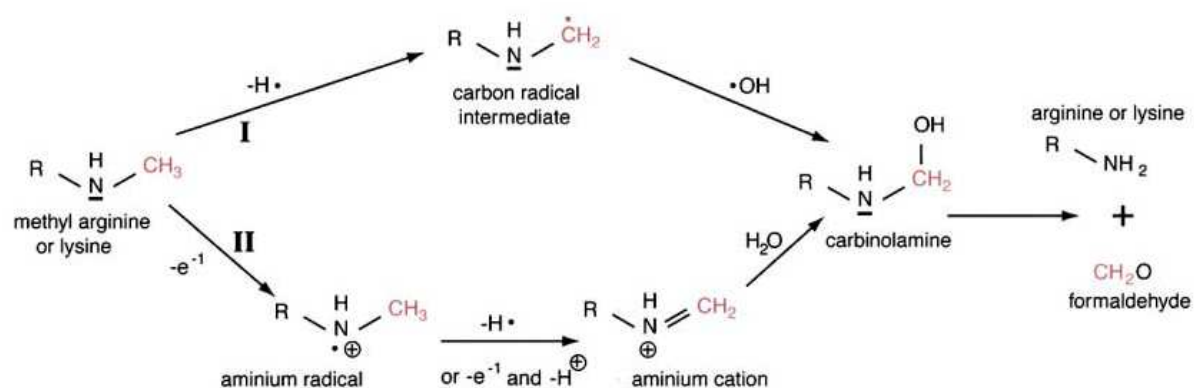


Figure 2. 6: Two possible reaction mechanisms for arginine and lysine demethylation [38].

The first pathway, given in Figure 2.6, involves single electron transfer, while the second one contains two electron transfers. At the first step of the first pathway, single electron transfer leads to the formation of carbon radical intermediate and then an $\cdot\text{OH}$ radical is added to form a carbinolamine. In the second pathway, an electron is transferred to the methyl lysine or arginine and an aminium radical is formed, which is then converted to an aminium cation via second electron transfer. The aminium cation is then converted to carbinol amine in the presence of water. The final step of both pathways is common, which involves the spontaneous structural rearrangement of carbinol amine that yields the demethylated lysine or arginine and the formaldehyde by product.

LSD1 is substrate specific, i.e. it can only demethylate mono- and dimethyl lysine residues, and this specificity is stemming from the chemical nature of the amine oxidation reaction. It is consistent with the observations done on other FAD-dependent amine oxidases, e.g. MAO, PAO [16]. In this light, if two pathways are analyzed, it is observed that the second pathway is more appropriate for LSD1. This is stemming from the fact that trimethylated lysine, having no free electron, cannot donate a single electron in order to be transferred in the first step (Figure 2.6). Therefore, the second pathway can only work for mono- and dimethyl lysines, which have a free electron pair on N-terminus. Based on the second pathway, Shi et al. proposed the reaction mechanism (pathway) for LSD1-catalyzed demethylation process given in Figure 2.7 [16]. In this mechanism, the first step essentially consists of two sub-steps involving two hydrogen transfers from lysine to FAD and thus it is considered to be the rate-determining step and catalyzed by LSD1. Here FAD is used as the oxidizing agent (being reduced to FADH₂). In contrast, O₂ re-oxidizes FADH₂ molecule to FAD (from hydroquinone

form to quinone form). Therefore, demethylation process is dependent on the presence of FAD and molecular oxygen.

A Postulated pathway for Demethylation of diMeK4H3 by LSD1

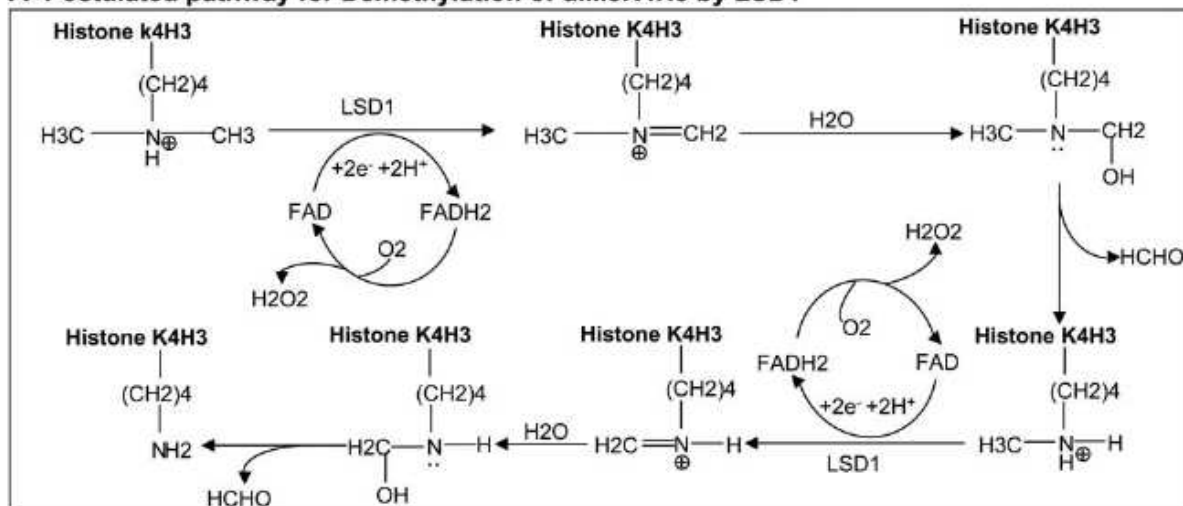


Figure 2. 7: Reaction Mechanism for LSD1 demethylation proposed by Shi et al. [16].

In order to be able to analyze the demethylation process at molecular level and create the energy profile of the mechanism, fine details of the mechanism should be known. The fine details can be considered as “which electron pairs attack to which atoms”, “which bonds are broken”, “which new bonds are formed”, “which atoms accept electrons while some others lose their electrons”, and etc. However, both of the mechanisms introduced by Bannister et al. and Shi et al. are overall mechanisms and do not contain the fine details of the mechanism. In order to fill in the blanks, a chemically-plausible and fine-detailed mechanism should be developed. For this purpose, some additional mechanism related to amine oxidation should also be considered. As LSD1 is a mono-amine oxidase, the modified radical mechanism, which was proposed by Silverman et al. long before the discovery of LSD1, may be analyzed to gain information on the amine oxidation that is carried out in the first step. The mechanism can be adapted to the demethylation mechanism, by replacing the amine substrate with the methylated lysine that is an amine by nature (please refer to Figure 2.8).

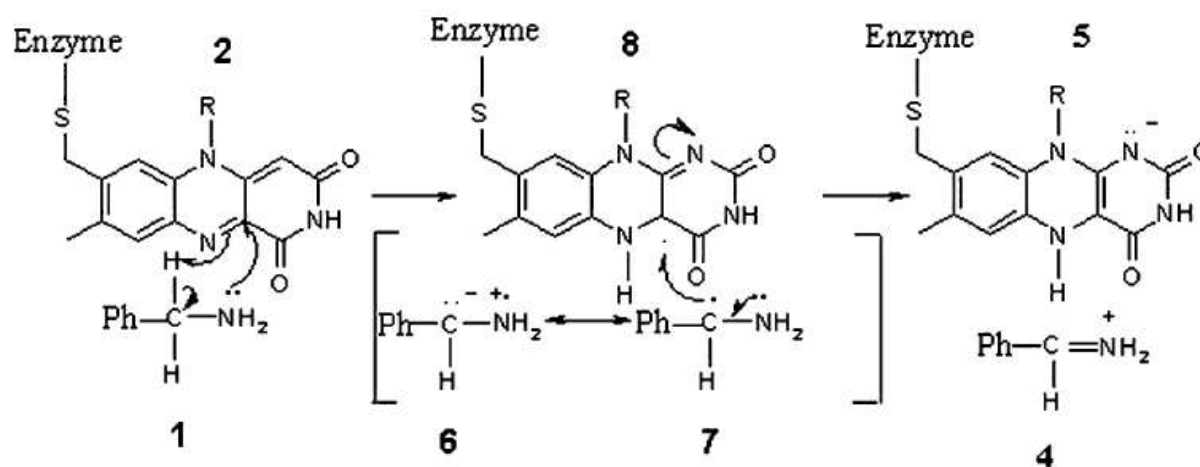


Figure 2. 8: Modified radical mechanism proposed by Silverman et al. [92].

There have been other mechanism proposals for amine oxidation besides the radical mechanism. There are four mechanisms: Hydride transfer mechanism [93, 94], addition-elimination mechanism (also known as polar nucleophilic attack or adduct forming) [95] and carbonion mechanism. The graphical representations of these mechanisms are given in Appendix A. Among these oxidation mechanisms, carbonion mechanism (see Appendix A1) is the most rooted one (proposed two decades ago for the first time) but not likely to occur because of the following. Carbonion mechanism proceeds by the abstraction of the α -proton by an active site base residue and that forms a carbonion [94]. However, this is not a favorable process; since α -proton is not acidic enough (pK_a value being greater than 30) and some X-Ray crystal structure studies showed that, at least for MAO-A and MAO-B, there is no active-site base that can abstract the α -proton [20]. This is likely to be the case for LSD1, as well.

Moreover, addition-elimination reaction (see Appendix A2) mechanism involves the attack of the amine nitrogen atom of the substrate to one of the electrophilic sites of the flavin, i.e. C2, C4a or N5. Afterwards, an active site base residue deprotonates and makes a β -elimination. However, this mechanism is opposed by Silverman et al. [96], who allege this mechanism may violate the microscopic reversibility. This mechanism was also opposed by Ralph et al. [97], who allege that this mechanism is not consistent with the previously observed ^{15}N isotope effect values, at least, for TMO (a mono-oxygenase) and it is likely to be the case for LSD1. Nevertheless, Yelekci et al. [20] used this polar nucleophilic mechanism to

study *in silico* the MAO-A reaction path and the effects of different substitutions of the side group of the amine substrate and obtained satisfactory results.

On the other hand, direct hydride transfer mechanism involves the direct proton transfer from C-terminus (or α -Carbon) of the amine substrate to N5 atom of the isoalloxazine ring of flavin. (See Appendix A3). Ralph et al. argue that the hydride transfer mechanism is more plausible for oxidation mediated by TMO based on their large ^{15}N isotope effect result that is consistent with irreversible CH bond cleavage [97]. However, the question that if these results are valid for other amine oxidases remain seeking further research [3]. In contrast, Yelekci et al. find the hydride transfer mechanism unlikely, since hydride transfer involves high-energy barrier [20].

Among these four oxidation mechanisms proposed by different groups, we chose to investigate the radical mechanism proposed by Silverman, since it is more compliant with the two-electron transfer mechanism proposed by Bannister et al. that is highly appropriate for LSD1. (The reasons are explained above). Hydride transfer also seems to be applicable for LSD1-catalyzed demethylation process, although it is not analyzed in the scope of this thesis. In this frame, however, a new chemically-plausible and logical reaction pathway should be prepared. We propose a pathway for direct hydride transfer mechanism that is modified for LSD1-catalyzed demethylation process. (See Appendix A4 and A5). This proposed mechanism may be directly used in reaction path analysis as future work.

Shi et al. were able to prove that the reaction products of the demethylation reaction catalyzed by LSD1 are demethylated lysine and formaldehyde. Those are consistent with the expected products, which are the same products as of other amine oxidase-mediated reactions [16].

Molecular oxygen, O_2 , is used to reoxidize the FADH_2 to FAD and it is important to know what is happening to H_2O_2 and formaldehyde after demethylation process, since they may have a deleterious effect on the promoters. We allege that after every two demethylation process two H_2O_2 molecules may interact and decompose to give molecular oxygen, O_2 , and two water molecules according to equation 2.1, that replace the consumed reactants in the demethylation process.



H_2O_2 has been proven to be functioning as a signaling molecule in a number of physiological events, such as apoptosis, cell cycle progression, cell differentiation and transcriptional regulation, and also in pathological processes, such as neural-degeneration [98]. At this point, the question remains to be answered if the hydrogen peroxide is a waste product or it performs specific biological function in chromatin remodeling [40].

On the other hand, the formaldehyde product of the demethylation mechanism can react with any amine *in vivo*, therefore it can be damaging to DNA within the cell nucleus and it should be cleared from the reaction pocket by the means of a possible cycle that recycles formaldehyde to S-adenosyl-L-methionine (SAM). SAM is used as a methyl group donor in methylation process. In this frame, the sustainability of the methylation/demethylation balance is provided [38]. Hypothetical formaldehyde cycle is given in Figure 2.9.

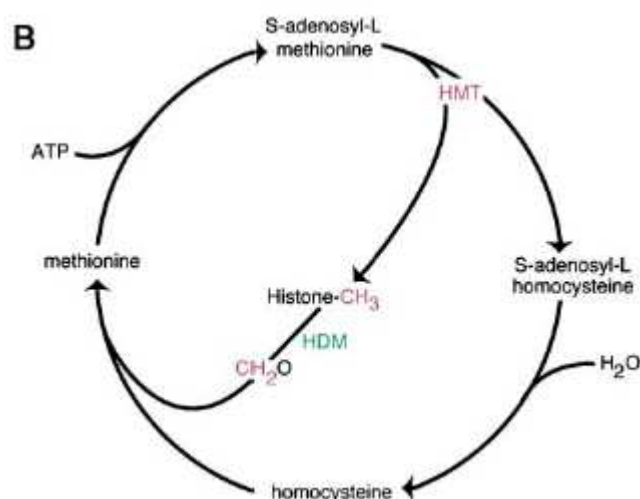


Figure 2. 9: Formaldehyde recycling scheme proposed by Thyiak et al. [99].

2.4.3.2 Information on Mechanisms Utilized by Jumonji Family Enzymes

JHDM family enzymes are members of Fe(II)/2-oxoglutarate-dependent dioxygenase (or equivalently hydroxylase) superfamily. The reaction pathway of JHDM family enzymes makes them able to demethylate trimethylated lysines in addition to di- and monomethylated lysines. As the reaction pathway, a quaternary complex, which consists of 2-oxoglutarate, Fe(II), and methylated lysine, reacts with molecular oxygen, O_2 . (See Appendix C2). The active site residues His188, Glu190, and His276 are conserved and have been shown to be involved in chelation of Fe(II) [17]. As the first step in reaction pathway, Fe(II) donates an

electron to molecular oxygen; as the result, a superoxide radical and Fe(III) are formed. The activated molecular oxygen attacks at the ketone carbon of 2-oxoglutarate and an Fe(IV) peroxyhemiketal bicyclic intermediate is formed. At the next step, the peroxyhemiketal bicyclic intermediate is carboxylated to form succinate, Fe(IV)-oxo intermediate and carbondioxide, CO₂. The methyl carbon of the lysine is oxidized by the Fe(IV)-oxo intermediate leading to the formation of a hemiaminal intermediate, which is common with the LSD-catalyzed demethylation mechanism (see Figure 2.7), and Fe(II) is regenerated. This hemiaminal intermediate is spontaneously decomposed to formaldehyde and demethylated lysine by structural rearrangement, as is in the LSD1-catalyzed mechanism [17].

On contrary to the LSD1 catalyzed mechanism, the reaction mechanism of JHDM family enzymes does not require a lone electron pair on the ϵ -nitrogen of the methylated lysine. This makes the JHDM enzymes available to demethylate the trimethyl lysines.

In the active site of JMJD2A, Gly170, Tyr177, Glu190, Ser288 and Asn290 create the methylammonium binding pocket via C-H...O hydrogen-bonding to the methyl group(s) of the methylated lysine [17]. When the lysine is trimethylated, the third methyl group is directed toward the Fe(II) in the complex, while other methyl groups are away from the Fe(II). This is considered to be the reason why monomethyl and dimethyl lysines have a lower demethylation rate compared to the trimethyl lysines [17].

2.5. Inhibitor (Drug) Molecules Suitable for De/methylation Balance

Designing molecules that can inhibit the activity of the enzymes that catalyze a PTM is important, since the abnormal levels of PTM causes diseases. The balance, which is spoilt in case of a disease, of a PTM type can be regulated again via inhibitors (or namely drug molecules).

For an inhibitor molecule to be in the clinical use against diseases, it must be target-specific (selective), i.e. it must inhibit only the desired modification site, and effective in inhibition of the activity of the corresponding modification. A good inhibitor candidate should also cause no or only a trace amount of adverse effects in humans.

2.5.1 Methylation Inhibitors

Compared to other PTM types, e.g. acetylation and phosphorylation; methylation inhibitor research is at the entry level. Extensive research has been conducted for acetylation and phosphorylation modifications. Many inhibitor (drug) molecules have been designed for acetylation and phosphorylation and already in the clinical use against cancer therapy [42]. On contrary, there is a wide gap in inhibitor research in methylation modification. Only few inhibitor proposals have been made for methylation and only a few is in the clinical use.

In the literature, some inhibitor molecules have been reported so far for methylation process, including DNA methylation and histone tail lysine methylation [42]. For example, some S-adenosylmethionin (AdoMet or SAM) homologues, e.g. S-adenosyl-homocysteine (SAH) or sinefungin can inhibit DNA and histone methylation. They are not good inhibitor candidates as they show no selectivity for either DNA methylation or histone methylation.

Two lysine methylation specific inhibitors were reported, as the result of random screening of the compound library. Chaetocin, a fungal myxotoxin, was shown to inhibit *drosophila melanogaster* histone methyltransferase, i.e. Su(var)3-9, but it showed low or no inhibitory activity against G9a, EZH2 or SET7/9 [42]. Chaetocin was also shown to demethylate human hSu(var)3-9, but it yielded cytotoxic effects. The other inhibitor candidate is BIX-01294 that was shown to inhibit G9a methyltransferase activity, but had no effect on SUV39H1 or PRMT1 [42]. It can inhibit H3K9 dimethylation, but no effect on H3K27 or H4K20.

There are also inhibitor candidates that have been proposed against arginine methylation activity. Detailed information about them will not be given here because of the scope of the thesis; however detailed information can be obtained from [42]. Besides, one inhibitor candidate, BIX-01338, was shown to exhibit dual inhibitor activity against lysine and arginine methyltransferase activity [42].

2.5.2 Demethylation Inhibitors

LSD1 is a part of transcriptional complex that are involved in tumorigenesis. In this light, LSD1 inhibitors, which may be considered as antitumor therapeutic agents, have been designed.

Since LSD1 shows 20% structural similarity to other amine oxidases, i.e. MAOs and PAOs, and shares high-level sequence identity with them, the inhibitor molecules designed

for these amine oxidases have also been tried against LSD1 [40]. Expectedly, not all of the amine oxidase inhibitors are efficient against LSD1, but some of them are very efficient. For example, biguanide and bisguanidine polyamine analogues can reactivate genes that are silenced during the development of colon cancer [100].

Another example is the tranlycypromine (also known as PCPA or trans-2-Phenylcyclopropylamine) that is originally MAO inhibitor and a drug target for neurological disorders, i.e. antidepressant. It can form an adduct with MAO and also with LSD1 by binding covalently to the flavin ring of the corresponding enzyme preventing it from binding to its substrate. Tranlycypromine forms different adducts while bound to LSD1 and MAO, given in Figure 2.10 [5, 101, 102]. The inhibitory effect of tranlycypromine on LSD1 is shown in [103]. There are two proposed pathways for the reaction between tranlycypromine and flavin ring, which utilize the Single Electron Transfer mechanism of Silverman et al. Two mechanisms are given in Appendix B1. The crystal structure of LSD1 and tranlycypromine complex was obtained recently by several groups using X-Ray crystallography [5, 104]. The crystal structure is useful for reaction path analysis of LSD1 catalyzed demethylation process inhibited by tranlycypromine.

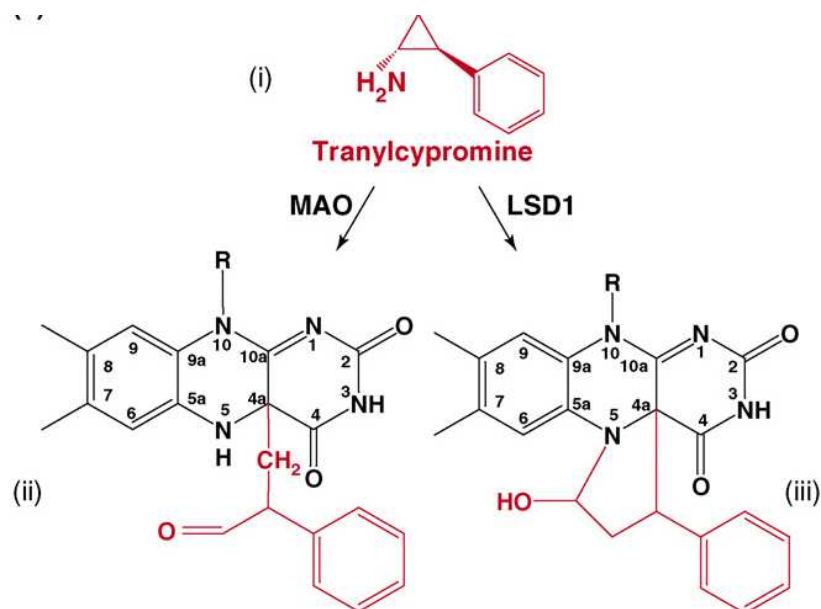


Figure 2. 10: (i) Molecular structure of tranlycypromine. (ii) and (iii) The adduct structures that tranlycypromine forms with the flavin ring of FAD that is either bound to LSD1 or MAO, respectively [40].

Based on the crystal structure of LSD1-CoREST complex, a histone peptide, which is modified by addition of a propargyl unit, a reactive chemical group, to the fourth lysine residue, can be used as an inhibitor molecule [105]. This modified lysine residue can covalently bind to LSD1 and the flavin ring of the FAD just like a regular histone does and acts as an inhibitor [40, 106]. The propargyl-added LSD1 and propargyl-added-LSD1-FAD complex are given in Figure 2.11. Some examples of propargyl-based type of inhibitors are selegiline and rasagiline, used as drugs against Parkinson disease [107].

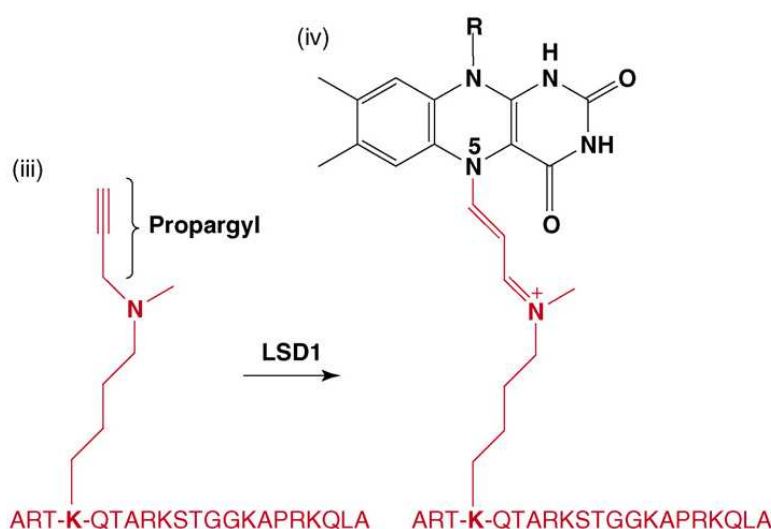


Figure 2. 11: (iii) Propargyl group added histone Lys4 residue. (iv) The adduct complex of Propargyl-added Lys4 and flavin ring of FAD [40].

As LSD1 is MAO homolog, inhibitors identified (or designed) for MAO can also be used for LSD1. In this frame, the study conducted by Yelekci et al. [108] presents a good comparison in terms of binding energy amongst the members of a collection of available inhibitor molecules designed for MAO-B. Among the 12 inhibitor candidates in the collection, the best candidate in terms of binding energy and inhibition constant (K_i) is (-)-(S)-1-Thiocarbamoyl-3-(4-methylphenyl)-5-(4-chlorophenyl)-4,5-dihydro- (1H)-pyrazole, which is given in Figure 2.12.

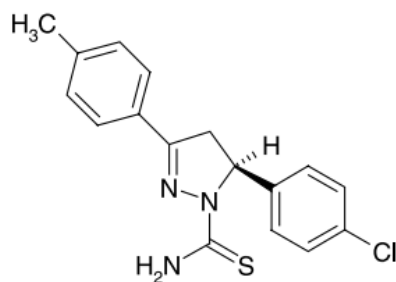


Figure 2. 12: The molecular structure of (-)-(S)-1-Thiocarbamoyl-3-(4-methylphenyl)-5-(4-chlorophenyl)-4,5-dihydro- (1H)-pyrazole [108].

On the other hand, also LSD2 was shown to be inhibited by tranlycypromine, as LSD1 and other MAOs are, using the same mechanics. Inhibition occurs when tranlycypromine is covalently bound to flavin ring [2]. Nonetheless, other LSD1 inhibitors, i.e. pargyline, deprenyl and rasagiline, were shown to have no effect on LSD2. These observations imply that tranlycypromine is a non-specific flavin-dependent amine oxidase inhibitor.

CHAPTER 3 COMPUTATIONAL METHODS and TOOLS

In this chapter, aim of this study, the theoretical methods, computational tools and the procedure that are utilized for achieving the aim are discussed in details. The results of the procedure followed will be given in the next chapter.

3.1 Aim of this study

In this thesis, reaction pathway analysis of LSD1-catalyzed histone H3K4 demethylation is carried out at molecular level using geometry optimization techniques implemented in *Gaussian09 (G09)* software package [109]. Geometry optimizations involve only quantum mechanics (QM) and hybrid Quantum Mechanics/Molecular Mechanics (QM/MM) methods, separately. Subsequent to geometry optimizations, vibrational analyses of the optimized geometries are done and Gibb's free energy and potential energy profiles corresponding to the proposed chemical mechanism are determined, as well as the structural properties of the stationary structures of the chemical mechanism.

3.2 Literature Review for Theoretical Methods and Computational Tools

Explaining the computational chemistry methods, referenced in the following sections, in sufficient detail, so that the reader gains all required knowledge, will be, of course, an inconclusive effort, since these are fruits of some very complex theories that have been developed by countless contributors for over a century. Nevertheless, the reader, who is interested in the theoretical and mathematical grounds of the methods, may refer to the books given in references [110, 111] and other valuable resources available in the literature.

In this light, in this subsection a very brief, general and non-comprehensive introduction is made to the concepts and computational methods that we made use of for this thesis. In this sense, this section lacks the specific details about the concepts, which will be gained from the corresponding references given in this chapter.

3.2.1 General Information

Chemical structures and reactions can be simulated numerically based on the physical laws, such as Quantum Mechanics or Molecular Mechanics (QM and MM). In addition to stationary chemical structures, unstable intermediate structures, e.g. transition state structures, can be modeled, which cannot be observed using today's available experimental methods.

Some common terms in computational chemistry being used in the thesis are as follows. The system of the atoms (or ions), whose thermodynamic properties and interactions with each other will be simulated using numerical methods, is called the simulation system. The spatial positions of the atoms in the system are together called the conformation. An optimized structure term stands for the conformation of the system that corresponds to the lowest energy in close proximity to the initial structure, which is given as input to the geometry optimization.

Given an initial structure (or equivalently conformation), the conformation that yields the (up to a certain extent) lowest energy can be determined by changing the conformation based on some chemical theories. This process is called geometry optimization. Geometry optimization proceeds stepwise and at each step, geometry (or conformation) of the chemical model is perturbed, i.e. position of atoms are changed, based on the forces acting on each atom. This stepwise geometry perturbation continues until the preset optimization thresholds are satisfied.

Geometry optimization can be easily explained on an energy surface plot. Energy surface is an N-dimensional plot of energy vs. change in the simulation system conformation. Here N is the degrees of freedom of the system, or namely, how many parameters are being changed controllably and energy is calculated against this change. As a simple example, for hydrogen molecule, H₂, energy surface can be plotted by changing the conformation of the hydrogen atoms found in the system and calculating the corresponding energy (see Figure 3.1). As there is only one degree of freedom of this system, which is the bond length of H-H bond, and only one parameter is changed at the same time, the energy surface is one-dimensional (1D).

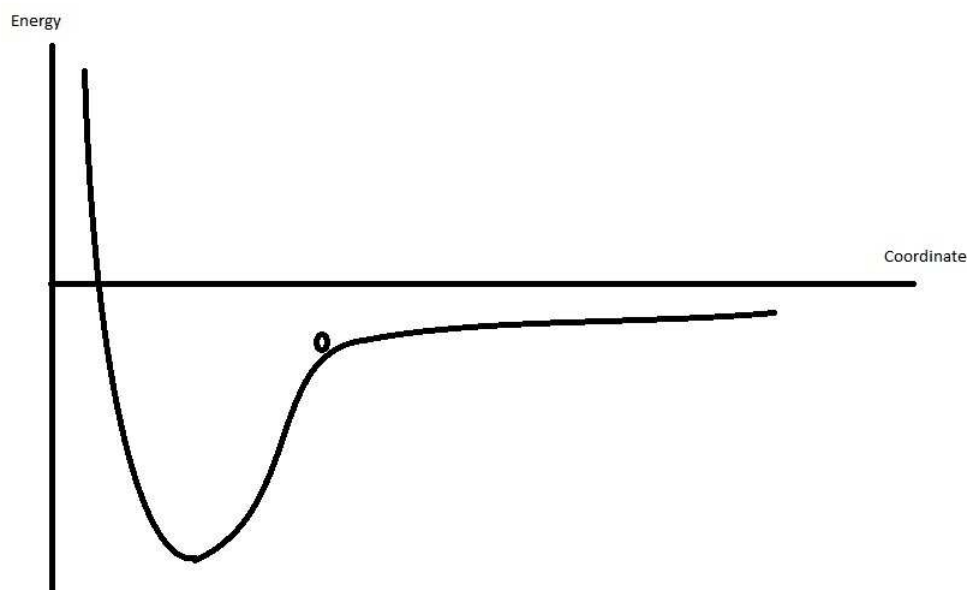


Figure 3. 1: Energy surface of hydrogen molecule, H₂.

During geometry optimization of H₂, the bond length of H-H bond is changed based on the forces acting on each H atom resulting from the electrostatic interaction between the H atoms, and energy is calculated. At this point, the analogy that the ball is trying to reach by making harmonic motions to the bottom of the well, i.e. the equilibrium or namely the lowest possible energy, is explanatory. Gravity, i.e. forces, pulls the ball, i.e. conformation of the simulation system, to the equilibrium.

In order to determine how far and in which direction to proceed on the energy surface at each step of the optimization, the derivative of energy with respect to coordinate, i.e. gradient, and forces are calculated. The second derivative of the energy with respect to coordinate gives the force constants, also known as Hessian matrix, which is used for finding the steepness of the step taken on the energy surface, thus is very useful in determining the next optimization step. In this frame, minima on the energy surface are determined as the points, whose gradient is zero, or below a predefined convergence threshold implemented in the optimization algorithm. Since the forces are negative of the gradient, at the minima points, forces are zero (or negligible), therefore they are also called stationary points. The system tends to stay at this stationary point unless an external effect is present.

The main purpose of the geometry optimization is to reach to the conformation, which is corresponding to the lowest possible energy, i.e. global minimum. This process is also

called minimization. When there is not only one minimum in the energy surface, like in the case of molecular hydrogen, there is the possibility of sticking to an undesired conformation, which is corresponding to a local minimum. In Figure 3.2, the points indicated with 1, 2, 3 and 4 are the energy minima. Point 1 stands for the global minimum by having the lowest energy compared to other minima. When initial conformation, which is given to geometry optimization as input conformation, is close to the point 3, the optimization algorithm will converge to point 3. If optimization algorithm is not specialized to come over the energy barrier between points 2 and 3, then it will not achieve to converge to the global minimum, i.e. point 1. In this scope, preparing initial conformation correctly with the desired structural properties is a crucial issue for avoiding undesired optimized geometries.

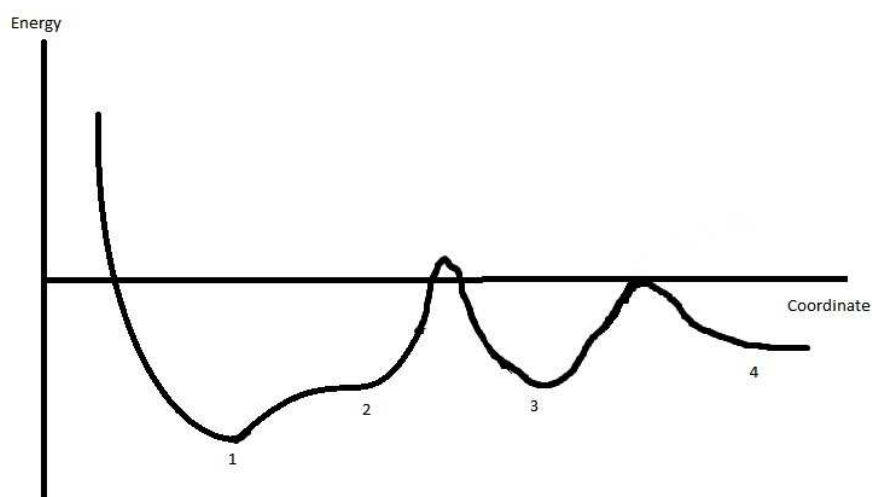


Figure 3. 2: An energy surface example

On the other hand, geometry optimization techniques can also be used to locate local and global maxima on the energy surface. This process is called maximization and maxima are characterized by the first derivative of its energy being zero with respect to the coordinate, as it is so also for minima. That means gradient is equal to zero. Maxima's second derivative of energy with respect to coordinate is negative, on contrary to the minima, which have positive values for the second derivative. Maxima generally act as a bridge between two local minima.

More importantly, energy profile of a chemical reaction can be created by combining the energy surfaces of the structures that is related to the reaction. These structures are

reactants, products and transition state(s). If we assume Figure 3.2 as an energy profile of a two-step endothermic reaction, then points 1 and 2 correspond to the reactants, while point 3 stands for intermediate products and point 4 depicts the products. The energy barriers between points 2 and 3, as well as points 3 and 4 depict the transition state structures of the first and second steps of the reaction.

Either QM or MM theories (or both) can be used for calculating the energies of the given chemical model structure throughout the geometry optimization. Energies are defined based on the electrostatic interaction between the nuclei and the electrons of the atoms placed in the given simulation system. Molecular mechanics theories use the laws of classical physics and make use of some tabulated force constants for electrostatic interactions and other thermodynamic parameters that are predefined for each atom type depending on which chemical moiety (molecule) it is placed in. In addition, the bond angles, bond lengths, dihedral angles, partial charges and etc are defined for each molecule type. All of these predefined values are packed into the packages called force fields. In these force fields, one atom type can be described by several atom types. For example, carbon atom is depicted in a force field by several carbon types based on being placed in a methyl group, carbonyl group, carboxyl group, and etc. In the literature, there are many force fields available, which are of general-purpose or specialized in a specific purpose, e.g. lipid, polar protein residues, and etc. In this study, Universal Force Field (UFF) will be used, which is a general-purpose force field and it will be explained in the upcoming parts.

Molecular mechanics (MM) methods do not explicitly treat the electrons. Instead, they use the interactions between the nuclei of the atoms. The effect of the electrons is implicitly integrated into the force fields via parameterization based on experimental results. As the electrons are not taken into account, MM methods are computationally inexpensive and used for large systems, i.e. having more than 150 atoms up to thousands of atoms. However, this approximation creates inaccurate calculations. Moreover, as MM methods do not treat electrons, they cannot model the chemical reactions, which involve forming and breaking bonds, i.e. electron transfer. Therefore, they cannot be used alone for simulating chemical reactions.

On the other hand, quantum mechanics (QM) involves solving the Schrödinger's Equation (Equation 3.1) in order to determine the energy and other related thermodynamics properties of a system.

$$H \Psi = E \Psi \quad (\text{Equation 3. 1})$$

In this frame, QM-based methods try to find the solution of the Schrödinger's Equation for a Hamiltonian (H), which is a mathematical operator that represents the sum of potential and kinetic energy of the electrons and nuclei of the simulation system [112]. Nonetheless, when the simulation system contains more than two atoms, the algebraic solution to Schrödinger's Equation becomes impossible. Therefore, some numerical methods have been developed for approximating the exact solution of the Schrödinger Equation for a simulation system. There are three main method types for approximating the Schrödinger's Equation, which are semi-empirical, *ab initio* and Density Functional Theory (DFT) methods. Each of the methods develops a different approach for the approximation [113].

Semi-empirical methods, e.g. Austin Model 1 (AM1), Modified Neglect of Diatomic Overlap (MINDO/3), Parameterized Model (PM3, PM4, PM6) and etc., use some experimentally obtained tabulated parameters for simplifying the computational complexity of the approximation to the Schrödinger's Equation. These methods can only be used for experimentally-well-defined systems and thus cannot be used for transition state structures. Therefore, these methods are used for large systems, as MM methods are. They are computationally more complex and able to provide more accurate comparison of two structures compared to MM methods. In this thesis, PM3 method is used for orientation of the molecules found in the simulation system based on one-on-one comparison of different conformations in order to obtain the best possible initial conformation for the geometry optimization at higher levels of theory [113].

Ab initio methods, e.g. Hartree-Fock (HF), Moller-Pleset (MP2 and MP4), Coupled Cluster Singles and Doubles (CCSD), Complete Active Space Self-Consistent Field (CASSCF) and etc, do not use any tabulated experimental parameters to approximate the Schrödinger's Equation, as *ab initio* stands for "from scratch". These methods form the linear equations for the approximation based only on the some physical constants, i.e. Planck's constant, speed of light masses and charges of atoms found in the simulation system. That

means, *ab initio* methods can be used for predicting energy of any type of system as it does not require any predefined parameter specific for the simulation system [113].

Density Functional Theory methods, e.g. Local Spin-density approximation (LSDA), Becke, Lee, Yang and Parr (BLYP), B3LYP, and etc, are relatively more novel compared to semi-empirical and *ab initio* methods. DFT methods have similar computational complexity of the least computationally complex *ab initio* method, i.e. Hartree-Fock. Nevertheless, DFT methods differ from the *ab initio* methods in terms of how they treat the electrons while approximating the Schrödinger's Equation. DFT methods take into account the effects of the electron correlation, i.e. how one electron react to the movement of another electron, as electrons do not want to be in the same spatial position according Pauli's exclusion principle. This electron correlation is calculated instantaneously, at every geometry optimization step, based on the interaction of two opposite spin electrons, while in HF, electron correlation is taken as an average. Therefore, DFT methods can provide the benefit of some higher order *ab initio* calculations at the computational cost of HF.

In addition, DFT methods yield more accurate, i.e. in close proximity to the experimental values, energy values and structural properties compared to semi-empirical and *ab initio* HF methods as they treat the electrons in the simulation system in a more realistic way [113]. Nevertheless, DFT methods lack some physical interactions, e.g. dispersion; therefore yield reaction energy barriers lower than the experimental values [18].

On the other hand, during simulation of a system, a basis set is used along with the model chemistry (or theoretical model) that is already explained above. A basis set is the numerical representation of the molecular orbitals of the simulation system, in which the electrons are placed. Basis sets define how much space the electrons will be confined in. In quantum chemistry laws, electrons have a finite probability of existing anywhere in the space in order to decrease the interaction with other electrons. Nevertheless, simulating infinite space is numerically impossible. Therefore, basis sets provides the simulating the effect of infinite space with less number of wave-functions to be solved numerically. In this frame, a larger basis set means more wave-functions are created for simulating the molecular orbitals and that leads to more accurate simulation and more computational cost [113]. Some basis set examples are (from smaller to larger) STO-3G, 3-21G, 6-31g, 6-311g, cc-pvQZ, lanL2DZ, SDD and etc. Each of the basis set is used for specific purposes, e.g. STO-3G is used for entry

geometry optimization, i.e. when the initial structure is far from the lowest energy structure, 6-31g and 6-311g for optimization of the organic molecules, other basis sets are used for the systems that contain inorganic components, e.g. metals, and etc.

Basis sets assign a group of basis functions, composed of linear combination of Gaussian primitives, to each atom in order to approximate its orbitals. The most basic (minimal) basis set is STO-3G (Slater-type Orbitals with 3 Gaussian primitives) and it provides 3 Gaussian primitives for forming the basis function for each atom. The size of the basis set is enlarged by increasing the number of primitives per atom via splitting the valance orbitals into different sizes. In this light, having more basis functions per atom increase the computation time, but offer more accurate simulation of the electron orbitals.

In addition, there are the polarized and diffusive functions, which are denoted with * and + terms. Polarized basis sets allow the orbitals to change shape by adding angular momentum. Only one * denotes only d orbitals (of heavy atoms) are polarized, while ** signifies p orbitals (of hydrogen atoms) are also polarized. On the other hand, diffuse functions combine s-type and p-type orbitals into one large orbital type, so that they span larger space and more accurate [113]. Another important point to regard in QM-based simulations is the open/closed shell differentiation (also known as restricted/unrestricted calculations). In closed-shell calculations, electron orbitals are forced to be filled by two coupled electrons, i.e. with opposite spins; while in open-shell calculations, orbitals filled by only single electron are allowed. This is done by creating two different types of orbitals, i.e. one for up-spin electrons, another for down-spin electrons, in open-shell calculations, whereas closed-shell calculations only involve one type of orbital. Besides closed-shell calculations, open-shell calculations are used in this thesis, besides closed-shell calculations, in order to simulate ions and radicals that are found in the simulation system [113].

Related to open/closed-shell calculations, spin multiplicity is a very important issue in QM-based calculations. It defines how the molecular orbitals will be filled with electrons and thus leading to open- or closed-shell calculations. Spin multiplicity is defined by the equation: $2S+1$, where S is the total spin of the simulation system. Each unpaired electron adds $\frac{1}{2}$ to the total spin. For example, a system with one unpaired electron is doublet, while systems with zero and two unpaired electrons have singlet and triplet multiplicities, respectively. Energy profiles are named after the multiplicities of the system. For example, energy profile obtained

with triplet spin multiplicity is called the triplet energy surface. Energy surfaces obtained with different multiplicity values essentially differ and they may intersect at some points called minimum energy crossing points (MECPs) passing from one surface to another. This spin-changing phenomenon sometimes accelerates the reaction with high energy barrier that cannot be overcome easily [114]. This case is explained well in Figure 3.3. In contrast, the spin-hopping phenomenon was introduced, which involves, even though MECP does not exist, the system changes its spin state, e.g. singlet, triplet, by jumping from one energy surface to another [115]. Furthermore, reactions occupying metals in their mechanisms are considered frequently to possess spin-changing action, while organic reactions are often considered to stick to single spin state.

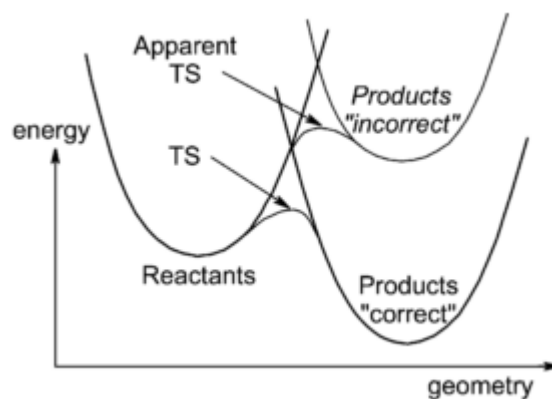


Figure 3. 3: Spin-changing phenomenon. If the spin state of the system is not changed after the transition state structure, the optimized structure of the products and the corresponding energy is “incorrect”.

3.2.2 Hybrid QM/MM calculations: ONIOM method

Enzymatic reactions, i.e. reactions catalyzed by an enzyme, are affected by the enzyme residues neighboring the species, which are directly involved in the chemical reaction. In other words, enzyme catalyzes the reaction by stabilizing the species reacting with each other, i.e. reactants, via intermolecular interactions, e.g. hydrogen bonding, repulsive, charge-charge, dipole and hydrophobic interactions. Reactants are placed usually in a multifaceted web of interactions with numerous neighboring enzyme residues. In order to simulate the system more realistically, i.e. precisely, by taking these interactions into consideration, the neighboring enzyme residues must be included in the simulation system in addition to the reactants [112].

Nonetheless, the computational cost of QM-based calculations increases drastically with increasing number of atoms building the simulation system. This drawback makes it unfeasible to model the reaction with enzyme effect at QM-level of theories, except semi-empirical methods, which provide faster but far less accurate simulation of the system compared to other QM-based methods [18]. In this frame, multi-level, i.e. hybrid, methods are required. Hybrid methods involve assigning a high-level (more accurate and computationally more complex, i.e. more time consuming) QM-based method to the part of the simulation system, which contains the species that are directly involved in the chemical reaction, and a low-level method (less accurate but more rapid) to the rest of the simulation system that contain enzyme residues and other species that are not directly involved in the reaction. In this case, the Hamiltonian, total energy of the system as defined in the first part, is given by Equation 3.2, in which H_{QM} , $H_{QM/MM}$ and H_{MM} denote the Hamiltonian of QM, QM/MM boundary and MM regions [112].

$$H=H_{QM}+H_{QM/MM}+H_{MM} \quad (\text{Equation 3. 2})$$

QM/MM hybrid approach was introduced in 1970s and there are several methods available in the literature for conducting QM/MM hybrid calculations [116]. Two major hybrid methods are empirical valance bond theory (EVB), and ONIOM (our Own N-layer Integrated molecular Orbital molecular Mechanics). Other methods differ from these methods in some minor details, such as treatment of the covalent interaction between two regions of the QM/MM system [117]. EVB method involves molecular dynamics (MD) simulations and Free Energy Perturbation (FEP) method for calculating the potential and free energy profiles of the simulation system. The algorithm details of EVB method is not given here as to the scope of the thesis; the reader interested in the details of the EVB method is advised to refer to [112, 118].

On the other hand, ONIOM method can treat a system by portioning it into desired number of layers, each of which may be treated with the same or different QM-based (it is also called MO based) or MM-based methods. That means all of the layers may be treated with the same method or different methods of choice. Generally number of layers is limited to two or three for convenience and also for preventing computational complexity. The electrostatic interaction between layers is handled with either classical (also known as

mechanical) embedding or electronic embedding methods compatible with ONIOM method [117].

In classical embedding, MM style point charges assigned to the atoms in the QM and MM region and the electrostatic interaction between these partial (or point) charges. On contrary, in electronic embedding, QM region atoms are assigned with QM-style charges and the partial charges of the atoms of MM region are included in Hamiltonian term of the Schrödinger's Equation. In this way, wave-functions describing the molecular orbitals respond to the changes in the MM part charge distribution by being polarized, so that a more accurate description of the electrostatic interaction between the QM and MM regions is provided [117].

Furthermore, for the treatment of the boundary atoms that are bonded to each other but belong to different layers, link atom approach [119] is implemented as complementary to ONIOM method on contrary to other QM/MM methods, which involve frozen orbital approach, e.g. Local Self-Consistent Field (LSCF) method [120] that yields slightly better performance, but not flexible for different systems, as this approach requires parameterization for every QM/MM method combination [117]. In contrast to frozen orbital method, link atom approach involves addition of imaginary hydrogen atoms to complete the valance electrons, which are missing, since the bonds are cut from the half at the boundaries.

As a part of the procedure followed for this thesis, two-layer ONIOM-based QM/MM calculations are done. In this light, how the ONIOM method handles a two-layer system will be explained using the ethane example originally given by Morokuma et al. [117]. In Figure 3.4, graphical QM/MM representation of ethane is given. The ethane is divided into two layers, i.e. QM and MM region, which are connected to each other via C-C covalent bond. In ONIOM, for describing the system, two subsystems are defined, i.e. real and model systems. The real system contains all of the atoms in the system and is treated only with the MM level theory. The model system contains the QM region atoms and it is handled at the QM and MM level, separately. The model system contains a link atom (LA) that is hydrogen, since otherwise the valance shell of link atom connection (LAC) will not be saturated. This link atom mimics the effect of the link atom host (LAH) that is placed in the MM region and not included in the model system. Here the link atom is chosen to be hydrogen, which is the default in ONIOM methods, but there is no constraint on the type of the link atom.

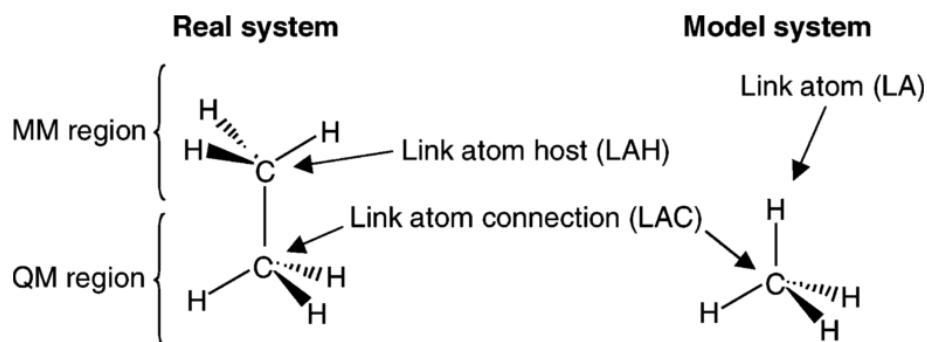


Figure 3. 4: QM/MM representation of ethane in ONIOM method [112].

In the frame of ONIOM, the total energy of a two-layer system is calculated using Equation 3.3.

$$E^{\text{ONIOM}} = E^{\text{Real:MM}} + E^{\text{Model:QM}} - E^{\text{Model:MM}} \quad (\text{Equation 3. 3})$$

In Equation 3.3, $E^{\text{Real:MM}}$, $E^{\text{Model:QM}}$ and $E^{\text{Model:MM}}$ terms refer to calculated energy of the real system at MM level, model system at QM level and model system at MM level, respectively. Moreover, the reader interested in the details of the ONIOM algorithm may refer to [117, 121].

3.2.3 Gaussian Software Package

Gaussian09 (G09) software package [109] is used for carrying out geometry optimizations and vibrational analysis, which are required for obtaining the energy profiles of lysine demethylation reactions within the scope of this thesis. *Gaussian* can perform the QM-based calculations using the following theoretical methods: Hartree-Fock, CIS, MP2, MP3, MP4 (SDQ), CID, CISD, CCD, CCSD, QCISD, BD, CASSCF, and all DFT methods, in addition to semi-empirical methods. In addition, following MM methods are implemented in G09: Amber, UFF, Dreiding. Among them, B3LYP (a DFT method), MP2 (an *ab initio*), PM3 (a semi-empirical method) and UFF will be used in the thesis.

Gaussian utilizes BERNY algorithm using GEDIIS [122] as the default geometry optimization algorithm for locating minima and maxima (TS structures). Optimization process starts by creating an initial guess for the second derivatives, i.e. Hessian, using the atomic coordinates and radii. Force constants are derived from Hessian matrix and the Hessian is improved based on the knowledge gained from the last optimization step. At each step of geometry optimization, wave-functions describing the molecular orbitals, in which electrons are occupied, are optimized using EDIIS and CDIIS methods [123], until it is converged to the threshold values, which are defined as 10^{-8} for RMS of all elements and the maximum

component value of electron density matrix, and 10^{-6} for the energy value. The wave-function optimization is based on the Self-Consistent Field (SCF) theory implemented in *Gaussian*. The fine details and operating procedure are described well in *Gaussian09 User's Reference* under keywords OPT and SCF. For online reference: [124].

On the other hand, *Gaussian* software package uses the following threshold values as the convergence criteria during the geometry optimization of a given system [113].

1. Forces acting on each atom must be lower than 4.5×10^{-4} .
2. Root mean square (RMS) of the forces must be lower than 3.0×10^{-4} .
3. The total calculated displacement of all atoms in the next geometry step must be lower than 1.8×10^{-3} .
4. The RMS of displacement components must be lower than 1.2×10^{-3} .
5. If forces are lower than 4.5×10^{-6} , then even if the displacement values are not below threshold values, the optimization is considered to be converged. This criterion is useful for large molecules, which generally have flat energy surfaces near to convergence.

OPT keyword, by default, creates an initial guess for SCF calculations using Harris guessing [125] at the beginning and use the guess from the previous step during optimization. At every geometry optimization step, *OPT* keyword calls the *SCF* routine for optimizing the wave-functions of electrons for given atomic positions.

At each geometry optimization step, atom nuclei are assumed to be fixed, according to Born-Oppenheimer Approximation, since their masses are ca. 1000-fold larger compared to electrons thus their vibration speed is almost zero compared to electrons. However, in reality nuclei of atoms (or ions) also do regular and predictable vibrations around the equilibrium state conformation. In this light, optimized conformation of the system is used in the calculation of vibrational frequencies of nuclei found in the system using *freq* keyword implemented in *Gaussian*. Frequency values of the vibrations are derived as the eigenvalues from the second derivative of the energy with respect to the coordinate. This calculation provides zero-point vibration energy (ZPE), thermal energy corrections, dipole moment, polarizability, as well as the thermodynamic properties of the system, e.g. total potential energy, Gibb's free energy, entropy, enthalpy, thermal energies, heat capacity and partition

function. This keyword provides vibrational normal modes, as well as IR and Raman spectra of the simulation system; however, this information is not used in the scope of the thesis. By default these values are calculated using the condition: 298.15 K and 1 Atm [113].

In thermodynamics calculations, the total energy of the system is calculated by summing up the calculated electronic, vibrational, rotational, and translational energies. ZPE and thermal energy corrections must be added to this total energy to obtain more accurate energy values. Here ZPE is the correction to the electronic energy of the molecule, which is calculated inaccurately because of the vibrations that do not exist at zero temperature in reality but persist in these calculations because of the approximation for solving Schrödinger's Equation. Likewise, thermal energy correction is required to take the effect of temperature and pressure in to the energy calculations [113].

Calculated vibrational frequency values and calculated ZPE/thermal correction values have profound systematic errors depending on the QM-based method and the basis set used. Therefore, they should be scaled using the corresponding scale factors determined in previous studies given in [113]. For example, the scale factors for MP2(Full)/6-31g(d) and B3LYP/6-31g(d) is given in Table 3.1, since these method/basis set combinations will be used in the thesis.

Method/Basis Set	Frequency	ZPE/Thermal Correction
MP2(Full)/6-31g(d)	0.9427	0.9646
B3LYP/6-31g(d)	0.9613	0.9804

Table 3. 1: Scale factors for some Method/Basis Set for correcting the vibrational frequency and ZPE/Thermal Correction.

Eigenvalues obtained from frequency calculations supply very valuable information on whether the optimized structure is a stationary point. If the structure being optimized is intended to be a minimum on the energy surface, then it must have no negative eigenvalue (or equivalently vibrational frequency). An optimized structure with negative frequency/ies means the minimization is not complete yet; it should be optimized further with tighter convergence criteria to locate a real minimum. This point corresponds to an N^{th} -order saddle point on the potential energy surface, where N is the number of negative eigenvalues. On contrary, if the structure is intended to be a transition state structure, then it must have one and only one negative frequency, which indicates an imaginary vibration. This imaginary

vibration should involve the movement of the atom(s), which must move from one point to another for the reaction to occur (usually involving bond breaking and formation). For example, in a hydride transfer reaction, one H atom is transferred from one species to another. The TS structure of this hydride transfer involves one imaginary vibration, in which the hydrogen atom moves between the two positions, where it is bound to each species, respectively.

Once a transition state is optimized and possesses only one negative frequency that corresponds to the normal mode connecting the desired reactants and products, it may be checked further if it is the correct TS structure via intrinsic reaction coordinate (IRC) [126] calculations. Using the *IRC* keyword implemented in *Gaussian*, starting from the saddle point, i.e. the desired TS structure in that case, the negative frequency normal mode is followed on the potential energy surface and two local minima conformations are determined, which are expected to be the desired reactant and products.

Freq keyword also leads to calculation of partial charges of each atom in the system that are then used in cooperation with the force constants in Gibb's free energy calculations of the system [112, 113]. Partial charges are found by distribution of the total charge of the simulation system using APT (Atomic Polar Tensors) [127] and Mulliken population [128] analysis methods. As the atomic charges are not a QM-observable variable, they are not unambiguously predictable from the first principles of thermodynamics. Therefore, the assigned atomic charges are, to some extent, arbitrary [113], however they may be used for getting a feeling on the charge distribution of the simulation system.

QM/MM calculations are also done with *Gaussian* software package, in which ONIOM method is used by default. *G09* allows users to combine any QM-based and MM-based methods for simulating a system. The boundary atoms are automatically defined and link-atom approach is also used. The details of the ONIOM method are explained in the previous subsection.

Potential energy and Gibb's free energy profiles, in addition to entropy and thermal energy profiles, for a chemical reaction, mechanism of which is proposed, can be obtained using QM-based calculation methods implemented in *Gaussian* software package. Using the profile, free energy barrier for the chemical reaction to start can be determined. The barrier of

the rate-determining step, i.e. the one with the highest energy barrier, is used in determining the rate of reaction by the means of Arrhenius Equation, which is given in Equation 3.4.

$$k = \frac{k_b T}{hc} e^{-\Delta G^* / RT} \quad (\text{Equation 3.4})$$

In Equation 3.4, k is the reaction rate, k_b , h , and R are the Boltzmann, Planck and Gas constants, whereas T , c and ΔG^* stand for temperature, concentration, which is taken as unity, and Gibbs free energy barrier for the rate-determining step (or equivalently free energy of activation), respectively. The reaction rate can also be determined experimentally, which makes it available to compare the theoretical reaction rates with the experimental ones.

Gaussian software package offers a calculation type, called with *Stable* keyword, which is used to determine if the optimized structure has a stable wave-function optimization. In this calculation type, it is checked if there is a possible other solution to molecular orbital wave-function equations corresponding to a lower energy compared to current solution. This is realized by removing the constraints applied on the wave-function, e.g. changing from closed-shell (restricted) to open-shell, changing multiplicity and etc. As the result, it is determined if the optimized structure (and wave-function corresponding to this structure) is indeed at the ground state as intended [113].

For example, when molecular oxygen, which is a component of the enzymatic system that will be investigated in this thesis, is analyzed with *Stable* keyword, it is observed that molecular oxygen should be treated at triplet surface. In details, when O_2 that is optimized at closed-shell singlet surface is analyzed with *Stable* keyword, a lower energy solution to the wave-function is found at the open-shell triplet surface. In this light, when we optimized the O_2 structure at triplet surface, it yielded lower energy value by 41.77 kcal/mol at B3LYP/6-31g* level compared to the singlet one. This fact that the molecular oxygen is more stable at triplet surface will be used in the calculations done in the scope of the thesis.

3.2.4 AutoDock Software Package

AutoDock [129] is an automated docking method for predicting orientation of macromolecule-ligand complexes. It creates random orientations of rigid macromolecule and the flexible ligand to be docked on the macromolecule and calculates the binding free energy of these conformations. The conformations with the same binding energy are clustered. The

lowest binding energy corresponds to the best binding orientation of the ligand. In this frame, among the clusters, the one with the lowest possible binding energy is selected and the average structure is calculated for this cluster.

Being flexible means that the ligand can make rotations and bending about its bonds and it can also make translations. The random orientations are prepared by the random walk of the flexible ligand on the specific region (a 3D grid) of the macromolecule specified by the user. The energy calculated at current step of the random walk is compared to the previous step's energy. If the energy increases then the current step is discarded. During random walking, hydrogen bonding plays a crucial role, as it strengthens the binding of the ligand to the macromolecule.

AutoDock 4.2 was chosen for docking process because of its wide-use and proven success (some example studies are: [108, 130, 131]). For searching and calculating the binding energy, four algorithms are implemented in *Autodock 4.2*: Local Search (LS), traditional Genetic Algorithm (GA), Lamarckian Genetic Algorithm (LGA) and Monte-Carlo-based Simulated Annealing (SA). Among these four algorithms, LGA yields the best binding (orientation) results based on the comparison study conducted by Morris et al. [131] with ca. 30 PDB files. The reader, who is interested in the details on how these algorithms work, is also suggested to refer to this article.

3.3 Computational procedure used in this study

In order to create the energy profiles, the following procedure is followed, which is given in numbers for ease to follow. These are the general procedures; they describe which simulation systems are considered and why they are chosen, as well as the results obtained from the calculations will be explained in Results and Discussion chapter.

The procedure described in this section was tested by reproducing the only-QM-simulation of polar nucleophilic mechanism of Mono Amine Oxidase (MAO) enzyme introduced by Yelekci et al. [20] at the beginning of the thesis study. This work was chosen, since LSD1 is homologous to MAO by having covalently-bound FAD residue and catalyzes amine oxidation. The calculation results, as indicated in the article, were reproduced with an accuracy of less than 1 kcal/mol with the corresponding QM-based methods.

All of the geometry optimizations, frequency calculations and single point energy calculations were done at gas phase. However, the enzymatic reactions occur in water

medium inside organisms. In this frame, the calculated energy barriers, atomic charges and other thermodynamic properties may be different from the real-life aqueous values. Therefore, as a future work, all of the reaction path analysis explained in Results and Discussion chapter may be repeated in the presence of solvent.

3.3.1 Preparation for the Geometry Optimizations

Before the geometry optimization can be done, some initial (input) structures, which reflect the image of the real-life interaction network between LSD1, FAD and H3-tail (enzyme, cofactor and substrate), should be prepared. For this purpose, two crystal structure files (chosen among 18 structures available in *Protein Data Bank (PDB)*) are used accessed via PDB IDs: 2H94 [66] and 2V1D [132]. 2V1D crystal structure contains LSD1, FAD, CoREST and the histone variant H3.1, whose fourth lysine residue is mutated to methionine. Nonetheless, this crystal structure lacks the crystal water molecules. Crystal water molecules, on contrary to other water molecule types, spend most of their time in specific spatial positions, so that they can be detected by X-Ray diffraction. In contrast, 2H94 crystal structure contains crystal water molecules in addition to LSD1 and FAD.

These two PDB crystal structures were combined, modified and cut using *Accelrys Discovery Studio (DS) Visualizer* software, so that it contains only 22 of all LSD1 residues, which are indicated to be forming the reaction pocket by Stavropoulos et al. [66]. These LSD1 residues affect the reaction through interactions with 16 H3 tail residues, FAD and crystal water molecules. The full list of the enzyme residues found in the system is as follows: Val317, Gly330, Ala331, Met332, Val333, Val334, Phe538, Asn540, Asp555, Asp556, Phe558, Glu559, Phe560, Leu659, Asn660, Lys661, Trp695, Ser749, Ser760, Tyr761, Tyr763, and His812.

After two PDB files are combined and modified, the fourth methionine residue of the H3 tail is mutated back to its original form (i.e. lysine) using built-in mutation tool of *PyMol* software and one or two methyl groups are added to the fourth residue and then this lysine residue is cleaned using *DS Visualizer's clean geometry* tool, which does a very basic geometry optimization. All of the system is not optimized using this tool, in order not to change the crystal structure of the system based on very low level optimization methods compared to the QM-based methods.

When the resulting structure is analyzed, it is seen that the molecular oxygen is missing in the structure. The orientation of the molecular oxygen is determined separately by docking using *Lamarckian Genetic Algorithm*, *Local Search*, *Simulated Annealing* and *Genetic Algorithm* implemented in *AutoDock 4.2* software package. Among these four methods, the lowest binding energy is obtained with Lamarckian Genetic Algorithm, as also indicated in reference [131]. The details of the docking procedure followed are given in Appendix E1.

At the next step, the residues in the system, except the one not defined in the force field, i.e. FAD, are checked for missing atoms in the PDB structure (because of the low resolution of X-Ray diffraction method) by comparing them with the OPLS-AA/L all-atom force field [133] using the *pdb2gmx* utility implemented in *GROMACS 4.0.5* software package [134]. Some residue types that need special care are treated efficiently using a user-interface and their structures are corrected according to user-choice by this utility. In more details, among the 38 residues that were treated with GROMACS, there are four lysine residues handled in protonated state (+1 charged), two arginine residues in protonated state (+1 charged), one histidine residue handled in H-on-ND1 state (also called HisA) (+1 charged), two aspartic acid residues in deprotonated state (-1 charged), and one glutamic acid in deprotonated state (-1 charged) making total formal charge of +4. The FAD and methylated lysine residues' structures are checked manually and modified according to corresponding chemical structures given in Figure 3.5.

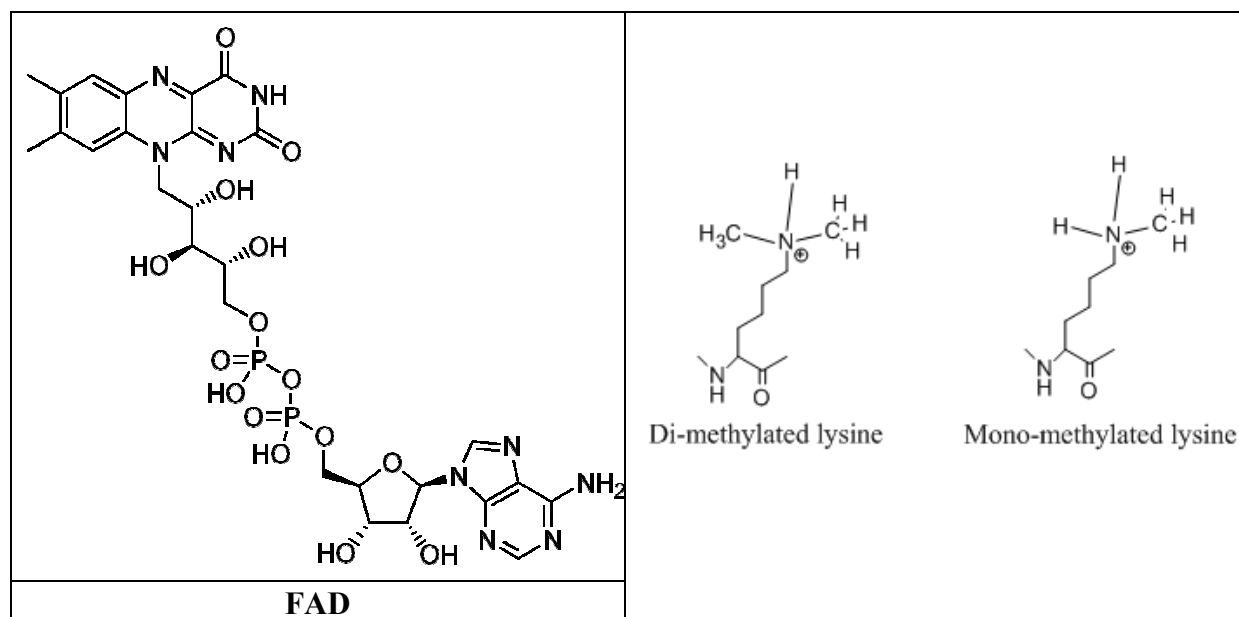


Figure 3. 5: Chemical structures of Flavin Adenine Dinucleotide (FAD), Mono- and Di-methylated lysines.

3.3.2 Geometry Optimizations, Vibrational Analysis and Single Point Energy Calculations using Gaussian09

3.3.2.1 Optimization Procedure

The input structure obtained using the procedure explained in the previous step is transferred from *DS Visualizer* to *GaussView* (visualization program of *G09* package) using *mol2* file format, which provides the least amount of loss in the connectivity information while transferring between two software programs. After the transfer, the connectivity of the system, i.e. the atom positions and the bonding between them, is checked again for possible adaptation problems. At this point, the system is ready for geometry optimizations. The whole structure is used for hybrid QM/MM calculations. Besides, if only QM-based calculations will be done, which treat only some part of the whole system, e.g. flavin ring of FAD, methylated lysine, molecular oxygen and water molecules, the rest of the structure except the desired components is deleted from the simulation system.

After a structure is ready to be given as input to geometry optimization, it is opted at B3LYP/6-31g* theory level via *OPT* keyword implemented in *G09* using default convergence criteria defined separately for wave-function and geometry optimizations. This theory level is chosen based on the observed success presented in the previous studies involving enzymatic reactions [19, 20, 135-137].

If the structure is optimized for a maximum on the energy surface (i.e. TS), *OPT=(TS,calcFc,noeigen)* is added on extra in the route line of *G09* input file, since TS calculations require approximate force constants at the start-up on contrary to minimizations. *Noeigen* option denotes that the test is turned off that is done for the presence of only one negative eigenvalue during optimization steps. This is required for large simulation systems. For both type of optimizations (minimization and maximization), *nosymm* option is added in order to prevent taking the symmetry of the system into account in the calculations for reducing computational complexity, since the system is not expected to be symmetric.

If the geometry and wave-function optimizations are having trouble in converging with the default criteria, there are some tricks, which are also applied in this study, given in Appendix E2.

In addition, semi-empirical PM3 level geometry optimizations are done for orientation and comparison purposes in the scope of the thesis (details of which will be conveyed in Results and Discussion chapter). That means the orientation of molecules inside the simulation system is tried to be determined by optimizing different manually-prepared input structures and comparing their energies for determining the one with the lowest energy. In that sense, trial-and-error-based orientation using PM3 resembles the docking procedure done with *AutoDock* software. The most probable orientation is further optimized at B3LYP/6-31g* level. Nevertheless, semi-empirical PM3 is used in this study only for locating energy minima, i.e. reactants or products, as this method cannot describe the transition state structures well, which often require open-shell calculations. In addition, although PM3 is very rapid compared to B3LYP method, the energy values it produces are less accurate and is only used for comparison purposes.

In hybrid QM/MM calculations, QM layer includes flavin ring of FAD residue, di- or monomethyl lysine (or its derivative as will be shown in the chemical mechanism that will be given in Results and Discussion chapter), crystal water molecules and molecular oxygen, MM layer consists of the tail part of the FAD residue (except flavin ring) and in some calculations also the neighboring enzyme residues. QM layer is handled with B3LYP/6-31g* level theory, while the MM layer is treated with Universal Force Field (UFF) [138], a molecular mechanics method. B3LYP/6-31g* and UFF are chosen for the QM/MM calculations because of their average computational complexity and satisfactory performance in simulating enzymatic systems as shown in the previous studies [1, 117, 139-151]. UFF as the molecular mechanics method is recommended for QM/MM calculations, since it yields reasonable energy barriers compared to experimental values, as shown in the previous studies, as well as our observations on the easy-to-use feature of UFF that requires no pre-configuration (or parameterization) before the use with different systems. In this work, we also tried AMBER [152] and Dreiding [153] force fields implemented in *G09*, but we countered almost no problem using UFF compared to other force fields. In contrast, atom type of each atom in the MM-layer must be defined according to UFF atoms types. This is carried out automatically by *GaussView*.

The effect of the charge of the MM layer on the electrostatic interaction of QM and MM layers are taken into consideration via mechanical embedding (by default in *G09*) during

QM/MM. The charge of the MM layer atoms are assigned to be zero by default, and the total charge of the molecules in MM layer can be distributed to each atom using *uff=qeq* option. This increases the accuracy of the calculations.

In QM/MM calculations, the input file is checked for the link atoms to determine if they are correctly assigned. *GaussView* assign automatically the link atoms in QM/MM calculations, but these link atom assignments must be checked manually for any false link atom assignment.

3.3.2.2 Frequency Calculations, Stability of the Geometry and High-Level Single Point Energy Calculations

After geometry optimization successfully completes, vibrational frequencies of the atoms around the optimized atomic positions are calculated using *Freq* keyword implemented in *G09*. This frequency calculation uses the atomic coordinates, Harris guess information and charge and multiplicity values from geometry optimization and calculates the force constants before the start-up. The resulting vibrational frequency values are analyzed for imaginary frequencies (negative eigenvalues) and other frequencies (positive eigenvalues). As it was explained in the previous subchapter, minimum structures must have no imaginary frequency, while TS structures must have one and only one negative frequency that corresponds to the vibration that connects the desired reactant and product structures. The calculated frequency values and the thermal corrections and ZPE values are scaled with the values given in Table 3.1.

Every optimized structure is further checked with *Stable* keyword for being ‘Stable’ by not having another wave-function solution corresponding to lower energy. The high-level single point energy calculations of the ‘stable’ structures is done at MP2/6-31g* level involving full-electron representation for obtaining more accurate energy values. MP2/6-31g* level calculations were shown in previous studies [1, 18, 19, 151] to yield highly accurate activation barriers, i.e. close to the experimental in 1 kcal/mol limit.

CHAPTER 4

RESULTS and DISCUSSION

The aim of this thesis study is to analyze the reaction path of the H3K4 demethylation process, which is catalyzed by LSD1, by obtaining the energy and entropy profiles of the reaction and analyzing the structural properties of the optimized geometries of the stationary points related to the reaction.

4.1 Proposing a Chemical Reaction Mechanism for Demethylation

In order to calculate energy profiles, a chemically-reliable and logical reaction mechanism should be proposed and used in analysis. This chemical mechanism shows the stationary point geometries and the structural changes during the reaction.

For this purpose, we start by giving a logical and chemically-consistent, at least to the best of our knowledge, chemical mechanism, which is given in Figure 4.1. This mechanism was created by us based on some former chemical mechanisms offered for LSD1 and amine oxidations, i.e. proposed by Shi et al. [16], Bannister et al. [38] and Silverman et al. [92], which are explained in Chapter 3. The mechanism is shown for dimethyl lysine, but it is also suitable for monomethyl lysine. The only change in the mechanism of monomethyl lysine demethylation is that a methyl (-CH₃) group of the dimethyl lysine is changed with -H group to replace the dimethyl lysine with monomethyl lysine.

The mechanism starts with a dimethyl lysine, oxidized flavin ring (in quinone form) of FAD, molecular oxygen and a crystal water molecule as reactants; and monomethylated lysine, formaldehyde and hydrogen peroxide are products. Whole reaction consists of four steps. In Figure 4.1, transition state structures are denoted with horizontal arrows, while minima on the energy surface, i.e. reactants and products of each step are also shown. Each of the stationary point structures is numbered for ease in identification; besides, special chemical name is given for the corresponding structure, if available.

The first step involves the hydride transfer from dimethylated lysine to quinone form of FAD. This step proceeds through a transition state, in which the hydrogen radical hangs between the nitrogen atoms of flavin ring of FAD and lysine. As products, this step forms the

aminium radical and FADH[•] radical. Step 2 involves the hydrogen transfer from lysine to flavin for the second time. At the end of this step, FAD becomes completely reduced to its hydroquinone form (FADH₂) and aminium cation is formed. In this light, FAD acts as a Lewis acid by accepting protons, while lysine acts as Lewis base by donating protons. After the Step 1 and 2, in an intermediary step, FADH₂ is re-oxidized to FAD by molecular oxygen yielding a hydrogen peroxide, H₂O₂, so that it can be used further in the next demethylation process.

In addition, the third step involves the attack of a crystal water molecule found in the reaction pocket to the aminium cation. As the result of hydration of aminium cation, carbinol amine is formed. This carbinol amine is spontaneously rearranged and separated into monomethyl lysine and formaldehyde. Step 3 and 4 are expected to be occurring spontaneously and therefore do not need the presence of the enzyme.

On the other hand, in the first step of LSD1-catalyzed lysine demethylation mechanism proposed by Shi et al. (Figure 2.7) is divided in to two separate steps in the mechanism proposed by us (Figure 4.1). The first step was indicated to be the rate-determining step, i.e. the step with the highest activation energy barrier, by Shi et al. Therefore, for determining which step corresponds to the rate-determining step in our mechanism, the activation barriers for two steps will be compared.

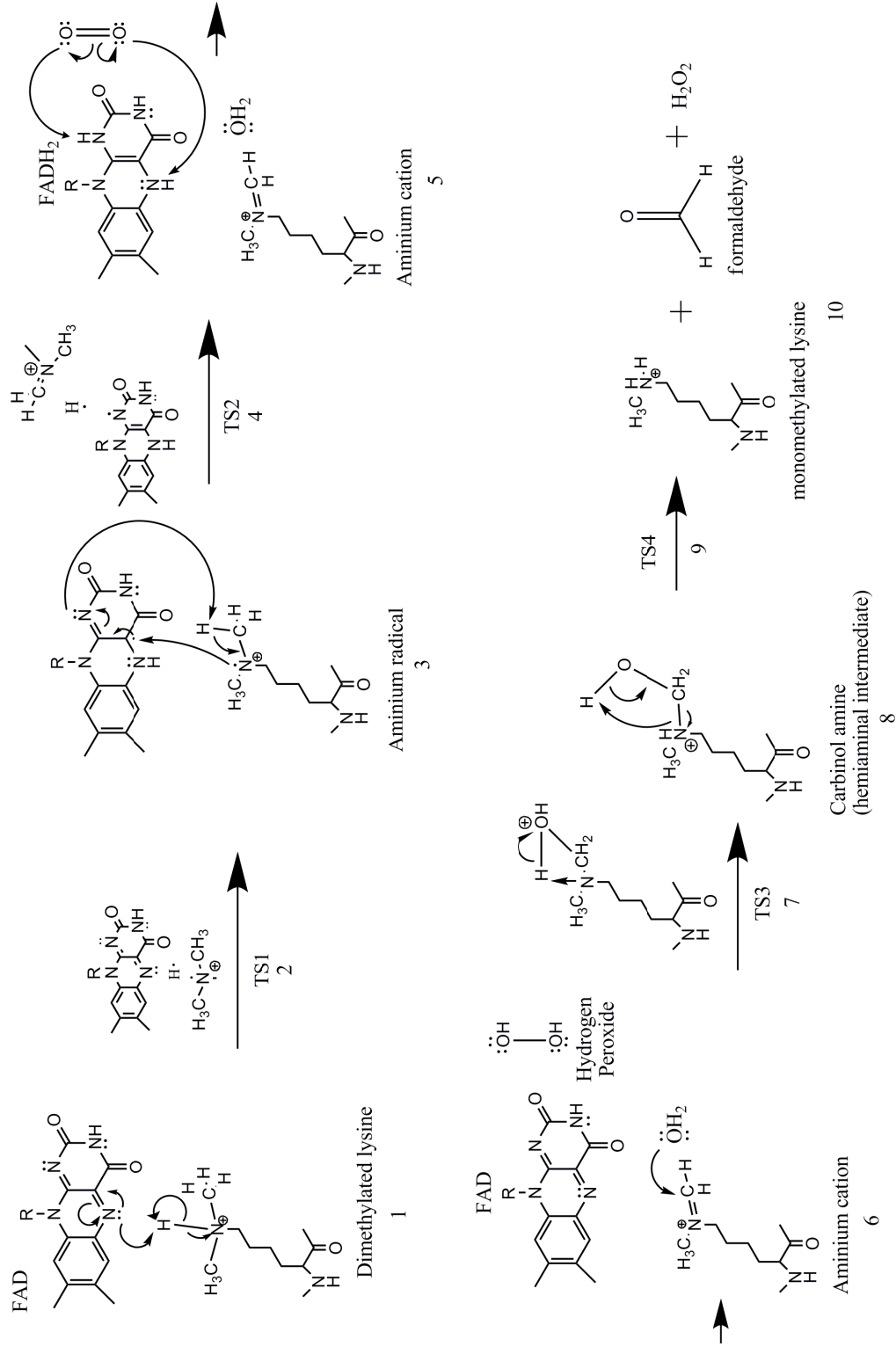


Figure 4. 1: Proposed chemical reaction mechanism for conversion of dimethyl lysine into monomethyl lysine via demethylation.

4.2 Only QM Simulation System and Calculation Results

In all only-QM systems, i.e. treated only with QM-methods, only the flavin ring of the FAD residue is included, since modeling the whole of the FAD molecule, which consists of 86 atoms, in addition to the lysine residue requires very high computational complexity.

We started with the only-QM system containing only dimethyl lysine and flavin. In the next step, we added one water molecule and then two water molecules to the system, respectively, and modeled the reaction path at B3LYP/6-31g* level until the end of the Step 2, since only this part of the mechanism requires the catalysis of the enzyme. As the last step, we added the molecular oxygen and analyzed the demethylation process of dimethyl lysine in the presence of flavin, molecular oxygen, and two crystal water molecules that are shown to be found in the reaction pocket (shown in the PDB file: 2H94). Our purpose was to determine the effect of each component of the reaction pocket on the reaction.

Finally, for only-QM calculations, we analyzed the monomethylated demethylation in the presence of two crystal water molecules, molecular oxygen and flavin and compared the observations with dimethyl lysine demethylation analysis.

Note: Color code of the atom types placed in the snapshots of the simulation systems is as follows. Carbon: Black, Nitrogen: Blue, Oxygen: Red, Hydrogen: White.

4.2.1 'Without-water' System

As the startup calculations, an only-QM system is created, which only consists of dimethylated lysine and flavin ring. This system will be referred to as 'without-water' system in the thesis. This system is modeled until the end of Step 2, as the Step 3 requires one crystal water molecule.

In the PDB file, 2V1D, H3K4 residue is mutated to methionine, so we mutated it back to lysine and pre-optimized with the *Clean Structure* tool of *Accelrys DS Visualizer* while preparing the structures. Therefore, the orientation of dimethyl lysine with respect to flavin is not known exactly and it may not be correct. Therefore, for guessing the orientation of flavin and dimethyl lysine, orientation procedure was carried out at PM3 level and also docking using *AutoDock* was applied.

In PM3 orientation process, 15 different randomly chosen input structures are optimized at PM3 level. Three models with the lowest energies are shown in Table 4.1. The

model yielding the lowest energy at PM3 level (Model 3) was further optimized with 6-31g* level.

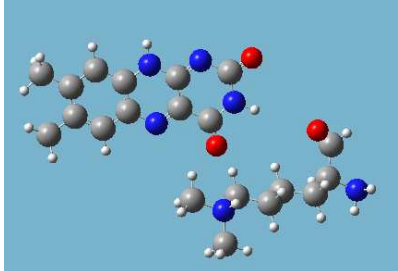
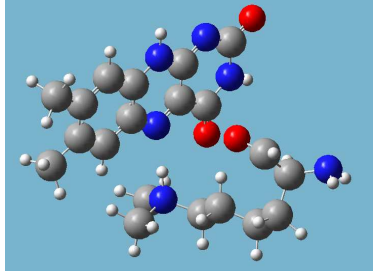

Model 1	Model 2	Model 3
		
E=0.08829au – 14.153D	E=0.08346au – 12.219D	E=0.07765au – 7.152D

Table 4. 1: Optimized structures of the six different models used for finding most probable configuration of the reactant of the first step (Dimethylated lysine and flavin of FAD). Optimizations done at PM3 level and total energy as well as the dipole moments of the systems are given [154].

On the other hand, the orientation of dimethyl lysine and flavin was tried to be determined using *AutoDock*. The binding energy values were found with different methods as follows. LGA: -2.34 kcal/mol, GA: -1.54kcal/mol, SA: -0.67, LS: -1.03. The best binding energy was obtained with LGA, as expected (referring to Chapter 3). The docked structure obtained with LGA is given in Figure 4.2. When the docked structure is compared with PM3 oriented structure, it is seen that the methylated N-terminus of the lysine is close to the N5 atom of the flavin ring in both structures. This orientation is obviously required for the reaction, i.e. Hydrogen transfer between the Nitrogen atoms, to occur.

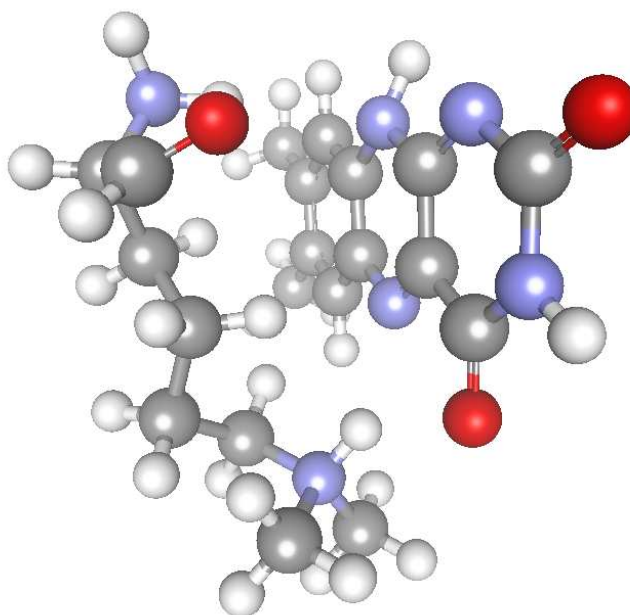


Figure 4. 2: The best orientation of dimethyl lysine and flavin obtained with AutoDock using Lamarckian GA. (Binding energy is -2.34 kcal/mol).

In order to compare the PM3 oriented structure (Table 4.1) with the one obtained with *AutoDock* (Figure 4.2), the optimized geometries obtained at B3LYP/6-31g* level are compared in terms of energy and dipole moment. The optimized geometry corresponding to PM3 orientation has a lower Gibb's free energy by 1.02 kcal/mol and this structure yields a dipole moment value of 43.79D, while the structure obtained with *AutoDock* yields 11.61 D. These results imply that both orientation methods produce different structures whose optimized geometries yield almost the same energy. *AutoDock* produces structures with less dipole moment, which corresponds to a more ordered distribution of the atomic charges.

For investigating whether giving two resonance structures as input to the geometry optimization makes any difference in the optimized structure, two resonance structures are prepared for optimizing the TS of the first step. The two proposals are given in Figure 4.3. After the optimization at B3LYP/6-31g* level, they converged to very similar structures with energy difference of lower than 0.01kcal/mol and a RMSD value of 0.0635. The optimized structures have very close negative frequency values of -934.26 cm^{-1} and 928.74 cm^{-1} [154]. That shows giving any resonance structure as input does not change the optimized geometry. Therefore, only one resonance structure is used for each geometry optimizations from that point on.

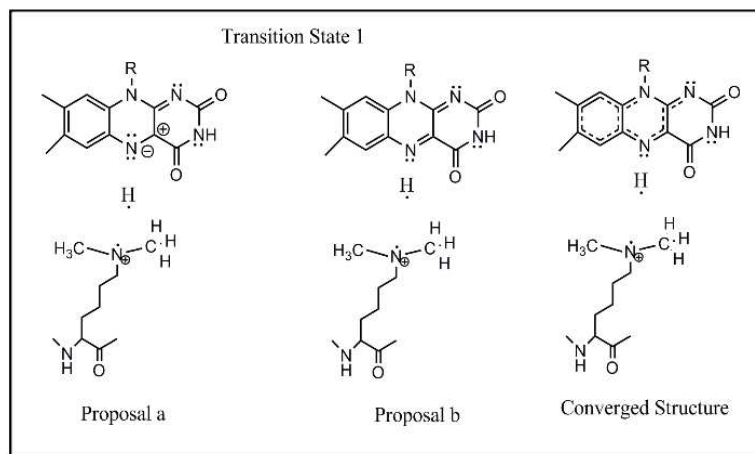


Figure 4. 3: Two proposals for geometry optimization of the transition state structure corresponding to Step 1 and the converged structure optimized at B3LYP/6-31g* level, +1 charged, TRIPLET [154].

For optimizing the TS structure of the second step, the input, which is given in Figure 4.4, is used and optimized at B3LYP/6-31g* level. The optimized structure is also given in this figure. It has only one negative frequency, -1024.24 cm^{-1} . On contrary to Step 1, minimum TS geometry for Step 2 could be obtained at PM3 level. The bond types between each atom of the optimized geometry are the same as the ones obtained at 6-31g* level, but the orientation of the two molecules differ compared to 6-31g* level structure. PM3 optimized TS2 structure also has 1 negative frequency vibration, which is at 1447.91 cm^{-1} and corresponds to the correct hydride transfer movement as in the 6-31g* level optimization. The RMSD between PM3 and 6-31g* optimized structures was calculated as 1.3458 [154].

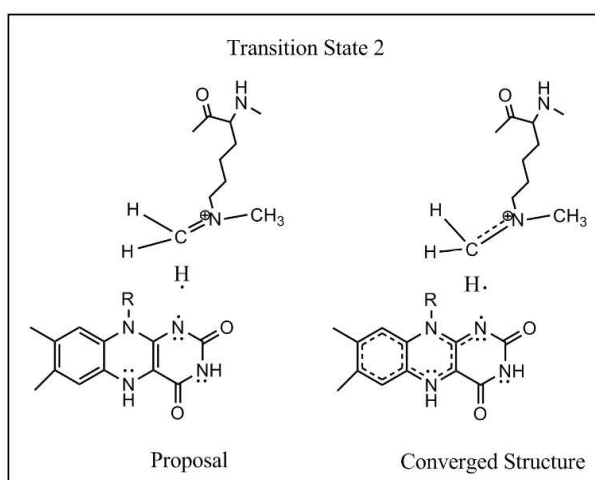


Figure 4. 4: Input and optimized geometry for TS structure corresponding to Step 2 at B3LYP/6-31g* level, +1 charged, TRIPLET [154].

In the literature, PM3 was shown to be unreliable in finding the energy profiles of chemical reactions by yielding high energy barriers. It is also known that, PM3 cannot model the transition state structures well, as they depend on experimentally predefined parameters. Since TS structures cannot be observed with experimental methods, TS structures cannot be modeled appropriately with PM3 method. To question this knowledge, potential energy profile for the demethylation reaction was tried to be obtained at PM3 level, given in Figure 4.5.

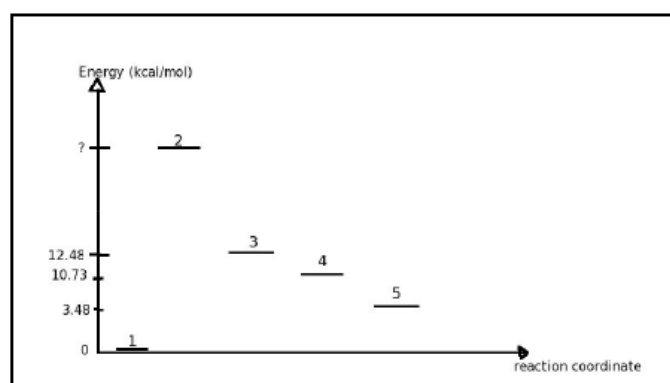


Figure 4. 5: Energy profile for demethylation of dimethyl lysine obtained at PM3 level [154].

In Figure 4.5, it is seen that the energy of the TS corresponding to Step 1 is absent, as convergence was not succeeded, even though many different inputs were tried. In contrast, the energy of the optimized geometry that was intended to be TS of Step 2 is lower than the reactants of Step 2. This is contradicting to what is expected, i.e. TS structures must have higher energy compared to reactants and products. Although PM3 yielded similar optimized geometry (having low RMSD value and close negative frequency values) compared to B3LYP/6-31g* for TS2 structure, these drawbacks prove that PM3 is insufficient for optimization of the TS structures in demethylation process. In this frame, PM3 have been used from this point on in this study only for orientation purposes.

For determining the orientation of aminium radical and flavin radical (FADH \cdot), PM3 orientation procedure is applied. The best orientation among 10 different proposals is given as the input structure to B3LYP/6-31g* level optimization. As it is shown in Table 4.2, aminium radical tends to shift to the top of flavin ring in order to minimize its interaction energy and to orient its methyl-group carbon atom with the nitrogen atom of the flavin, which will accept the hydrogen atom in the next step (i.e. Step 2).

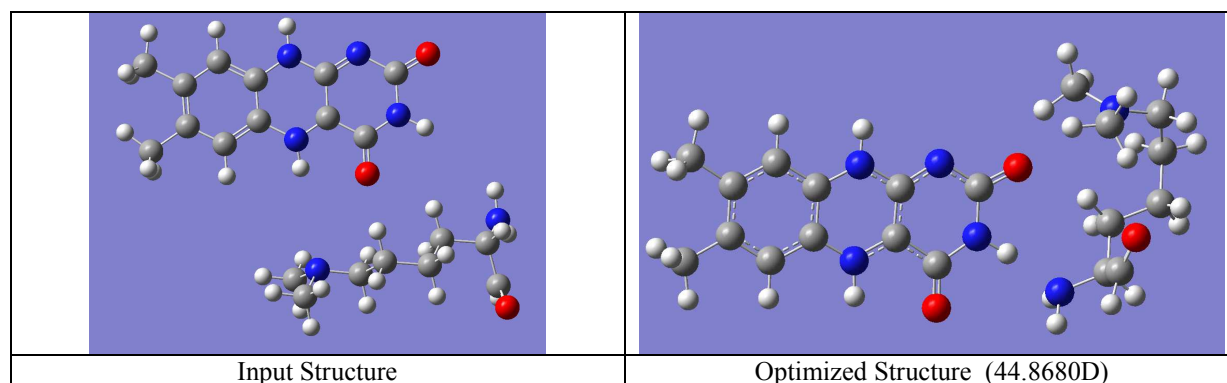


Table 4. 2: The best orientation of aminium radical and FADH obtained with PM3 orientation procedure. The optimized structure obtained at B3LYP/6-31g* level is given on the right hand side. Dipole Moment of the optimized structure is given in parenthesis.

As already explained previously, for computational complexity purposes, FAD molecule is truncated and only flavin ring part is used in only-QM calculations. In order to investigate the effect of truncating the FAD molecule on the reaction, we changed the R group of flavin ring, which was hydrogen in the previous case, into a methyl group, i.e. CH₃. Intention of placing a methyl group instead of hydrogen was to mimic the rest of FAD molecule, at least, to a small extent. The previous system (explained up to this point), whose R group is H, is called ‘Without-water’ system for short and the system, whose R group is CH₃, is called ‘R=CH₃’ for short in Figure 4.6. Figure 4.6 shows the potential and Gibb’s free energy profiles for these two systems in a comparative way. Energy profiles were obtained at B3LYP/6-31g* level and the structures corresponding to the numbers (i.e. 1 to 5) are given in Figure 4.1 (i.e. Structure 1 to 5). For obtaining the energy profiles, the structures are optimized at the following energy surfaces: Structure 1 (Step 1 reactants): singlet, Structure 2 (TS1): triplet, Structure 3 (Step 1 products, Step 2 reactants): triplet, Structure 4 (TS2): triplet, Structure 5 (Step 2 products): singlet.

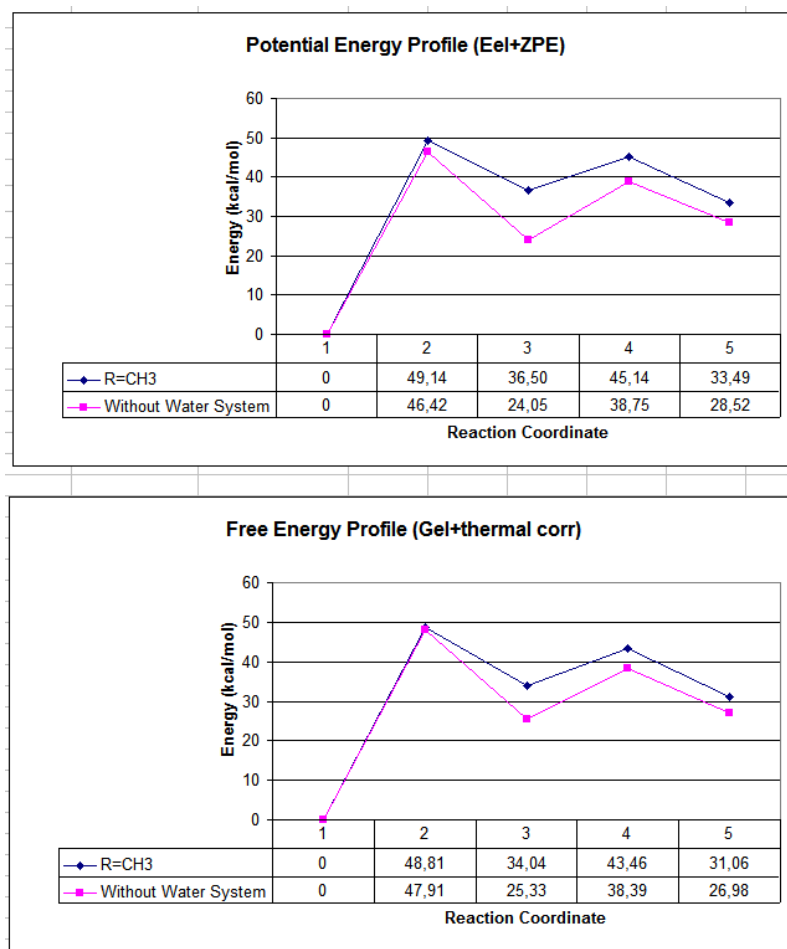


Figure 4. 6: Potential and Gibbs Free Energy Profiles of ‘R=CH3’ and ‘Without-water’ systems obtained at B3LYP/6-31g* level. Definitions of these systems can be found in text [154].

When Figure 4.6 is reviewed, it is seen that the ‘Without-water’ system has a free energy barrier of 47.91 kcal/mol and 13.06 kcal/mol for Step 1 and Step 2, respectively. The activation energy of the first step is so high that it cannot occur spontaneously and require enzyme activity. In contrast, the second step has a lower activation energy compared to the first step. These activation energy results allege that the rate-determining step of the reaction is the first step.

On the other hand, the other system, i.e. ‘R=CH3’, has energy barrier values of 48.81 kcal/mol and 9.42 kcal/mol for Step 1 and Step 2, respectively. When compared to ‘Without-water’ system, ‘R=CH3’ has a higher energy barrier for the first step, but lower energy barrier for the second step. This fact indicates that the effect of the rest of the FAD molecule cannot be interpreted by using only CH₃ to imitate the whole FAD molecule. The whole of the rest of the FAD molecule must be modeled, and it will be presented in the QM/MM calculations part.

Using the activation energies of the rate-determining steps obtained from ‘Without-water’ system and ‘R=CH₃’ system, the reaction rates are calculated via Arrhenius Equation (where temperature is taken as 25⁰C), as $9.86 \times 10^4 \text{ min}^{-1}$ and $4.61 \times 10^7 \text{ min}^{-1}$, respectively. The reaction rates are very low compared to the experimental value (i.e. 8.10 ± 0.20), because of the high energy barrier.

The calculated negative frequencies of TS1 (i.e. TS structure of Step 1) and TS2 of ‘R=CH₃’ system are -927.31 cm^{-1} and -900.96 cm^{-1} , respectively. Comparing these to the values of ‘Without-water’ system (928.74 cm^{-1} and -1024.24 cm^{-1}) [154], they are matching to great extent. This observation is consistent with what is expected, since both systems are making hydride transfer reaction having the same characteristics. Nevertheless, the negative frequency values are not equal for both systems, since the having different R groups is changing the characteristics of the optimized structure.

Moreover, we wondered if the changing the spin state of the system affects the energy profiles. For this purpose, we optimized Structure 1 and 5 at triplet surface rather than singlet surface. In contrast, Structure 2, 3 and 4 were already optimized at triplet surface, and optimizing them at singlet surface did not succeed despite our efforts, since they possess unpaired free electrons and cannot be optimized with closed-shell calculations. In Figure 4.7 are given the energy profiles for the triplet surface calculated for ‘Without-water’ system.

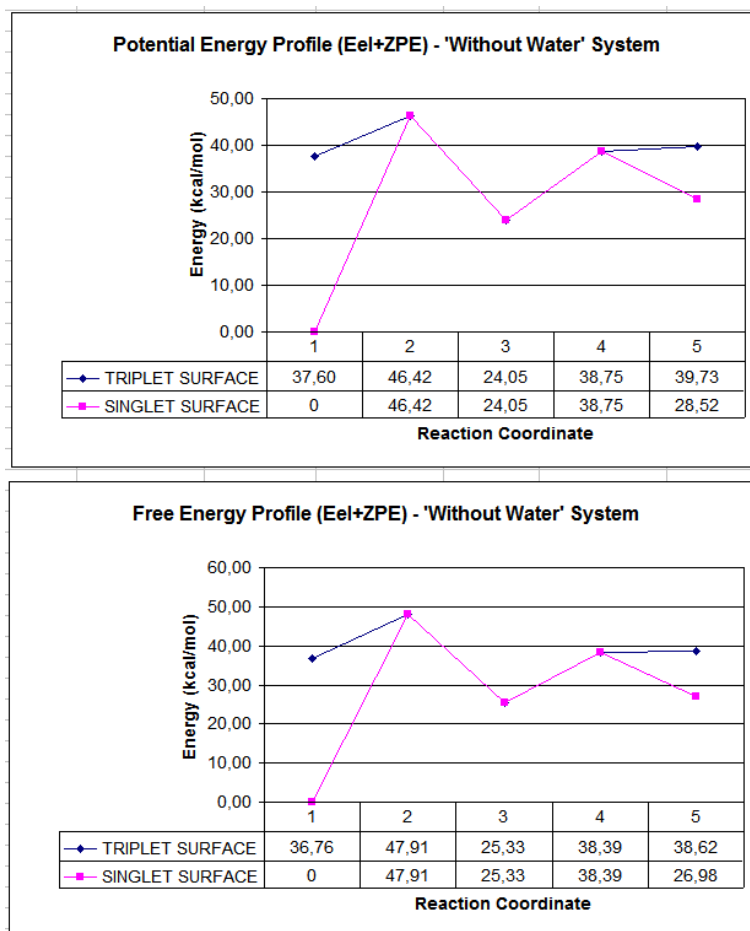


Figure 4. 7: Singlet and Triplet surface Potential and Gibbs's free energy profiles of 'Without-water' system (see text for definition). Profiles were obtained at B3LYP/6-31g* level. The energy values are adjusted with respect to the energy of Structure 1 at singlet surface.

In Figure 4.7, only Structure 1 and 5 energies differ on singlet and triplet surfaces since Structure 2, 3 and 4 could be optimized only at triplet surface. The potential and free energy of optimized Structure 1 is ca. 37 kcal/mol higher at triplet surface compared to singlet surface. This leads to a decrease in the activation energy of the first step (to a value of 10.15 kcal/mol). Therefore, it can be inferred that the first two steps must occur at the triplet surface in order to avoid the very high energy barrier (i.e. 46.42 kcal/mol). In this frame, it is likely to be correct to infer that the enzyme assures the reaction occurs at triplet surface during the first two steps.

On contrary, the second step has an energy barrier of 13.06 kcal/mol, which is higher than the first step (10.15 kcal/mol). In this light, the rate-determining step is the second step, rather than the first step at the triplet surface.

In contrast, at the end of Step 2 (Structure 5), singlet surface is more stable by providing an optimized geometry corresponding to less energy (i.e. 11.64 kcal/mol) compared to triplet surface. This fact may be interpreted as at the end of Step 2, the effect of LSD1 on the reaction is coming to an end the need for the enzyme presence vanishes. To sum up, the reaction is hopping to triplet surface at the beginning and it is coming back to singlet surface at the end of Step 2 by passing through an MECP.

4.2.2 Single-Water and Two-Water Systems

4.2.2.1 'Single-water' system

As the next step, an only-QM system is prepared with one water molecule in addition to dimethyl lysine and flavin ring. This system will be called 'Single-water' system from this point on in the thesis. Obviously, in the real case, there are two crystal waters in the reaction pocket and placing a single water molecule into the reaction pocket may not explain well the preferences of the system. However, it will give a clue.

The more favorable position of one water molecule with respect to other molecules was tried to be determined via PM3 orientation. Three different possible trials were prepared based on the position of the two crystal water molecules found in the reaction pocket according to the crystal X-Ray structure (PDB ID: 2H94). Positions of these two crystal waters in crystal structure are beneath the flavin ring and at the right-hand side of the flavin ring. (Their positions can be seen from Figure 4.10.) The optimized geometries of the three trials are given in Table 4.3. Characteristics of these trials are as follows: Water molecule is between flavin and lysine in trial 1, to the right-hand side of flavin in trial 2, and beneath the flavin and lysine in trial 3. Based on the energy values of PM3-level-optimized structures, the single water molecule prefers to be located beneath the flavin rather than right-hand side of the flavin. Besides, it does not tend to be placed between nitrogen atoms of flavin and lysine; instead it prefers to be beneath the lysine, where it can make hydrogen-bonds with lysine atoms.

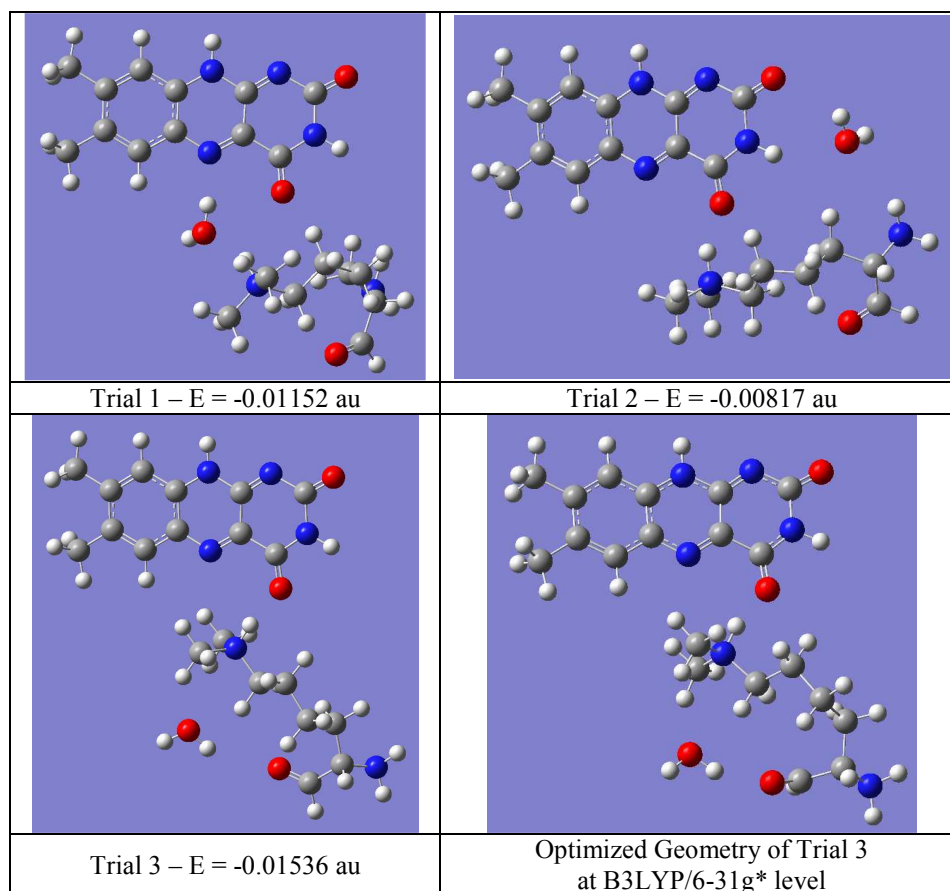


Table 4. 3: Optimized geometries of different orientation trials for ‘Single-water’ system (see text for the definition) obtained at PM3 level. Trial 3 is also optimized at B3LYP/6-31g* level.

In order to have an understanding on how a single water molecule affects the course of the reaction, energy profiles of the system were created at B3LYP/6-31g* level. In Figure 4.8, it is seen that the single water molecule does not change the progress of the reaction to great extent. It increases the energy barrier of the first step by ca. 2 kcal/mol, while it lowers the energy barrier of the second step by ca. 2 kcal/mol. In the overall picture, optimized energy of the structures is higher for ‘Single-water’ system.

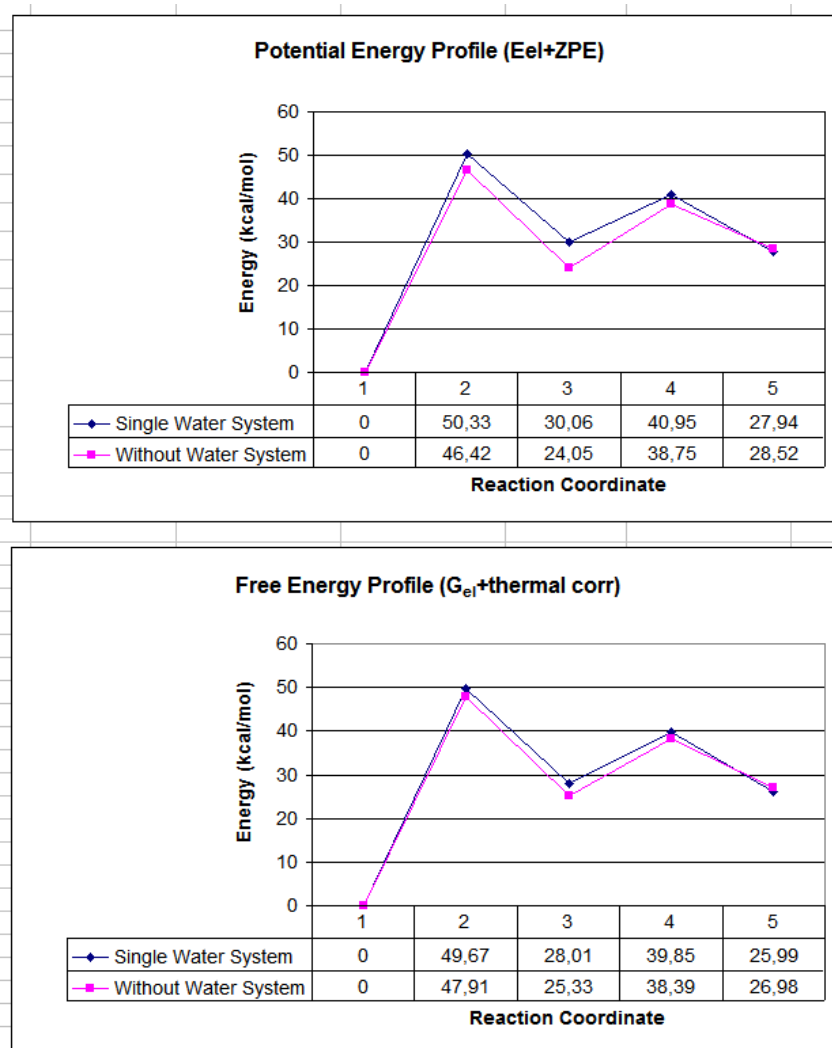


Figure 4. 8: Potential and Gibbs free energy profiles for ‘Without-water’ and ‘Single-water’ system (see text for definitions). Obtained at B3LYP/6-31g* level [154].

4.2.2.2 ‘Two-Water’ System

As the next step, one more water molecule was added to the only-QM system. For clearing ambiguity, this system contains flavin ring, dimethyl lysine and two water molecules. Besides, this system will be called ‘Two-water’ system, from this point on in the thesis.

First the orientation of lysine in the presence of two water molecule was tried to be determined. As the crystal structure (PDB ID: 2H94) shows the position of two reaction-pocket water molecules with respect to the flavin ring, this information was used in docking procedure. Lysine was docked on to ligand, which is consisting of FAD, two crystal water molecules and neighboring residues. Using *LGA* method, a position for lysine corresponding

to -2.12 kcal/mol binding energy was determined. This geometry was further optimized at B3LYP/6-31g* level. In the optimized structure (Table 4.4), lysine is pushed away by the water molecule placed beneath flavin ring to a position away from reacting position. That means, its positive-charged nitrogen atom is so far away from the nitrogen atom of flavin ring, and in this situation reaction is not likely to occur.

In this frame, in order to prepare the ‘Two-water’ system, we modified the optimized geometry we prepared for ‘Without-water’ system by adding two water molecules near to the positions shown in the crystal structure. Nevertheless, we had to pull the water molecule, by small extent, to the left-hand side of lysine, so that the lysine’s nitrogen atom could be placed near to flavin’s nitrogen atom. During the geometry optimization, the water molecule shifts beneath lysine in order to increase stability through hydrogen-bonds. Please refer to Table 4.4 for B3LYP/6-31g* level-optimized structures. The manually-created structure has a lower energy by ca. 2 kcal/mol compared to the structure generated by *AutoDock*. The structure created by *AutoDock* has a more equal distribution of charges compared to the manually-created structure, so that it has a very low dipole moment (9.23D vs. 49.66D).

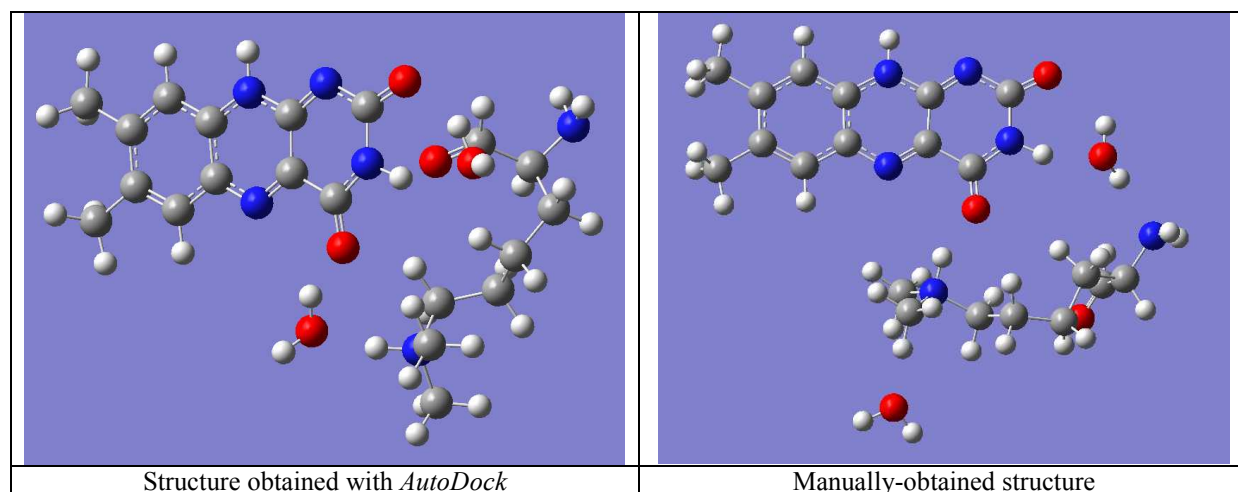


Table 4. 4: Two different Step 1 reactant structures of ‘Two-water’ system obtained manually and with *AutoDock*. See text for definition of the simulation system and methods. Both structures are optimized at B3LYP/6-31g* level.

On the other hand, the energy profiles of ‘Two-water’ system were created for singlet and triplet surfaces of ‘Two-water’ system at B3LYP/6-31g* level. Profiles are given in Figure 4.9. While preparing the singlet surface profiles, the triplet-surface energy values of structure 2, 3 and 4 are used, since these structures involve radicals (i.e. unpaired electrons) and could not be optimized with default convergence criteria at singlet-surface despite our

efforts. This is the same case as was observed for ‘Without-water’ system. In this frame, only Structure 1 and 5 have two different energy values for singlet and triplet surfaces.

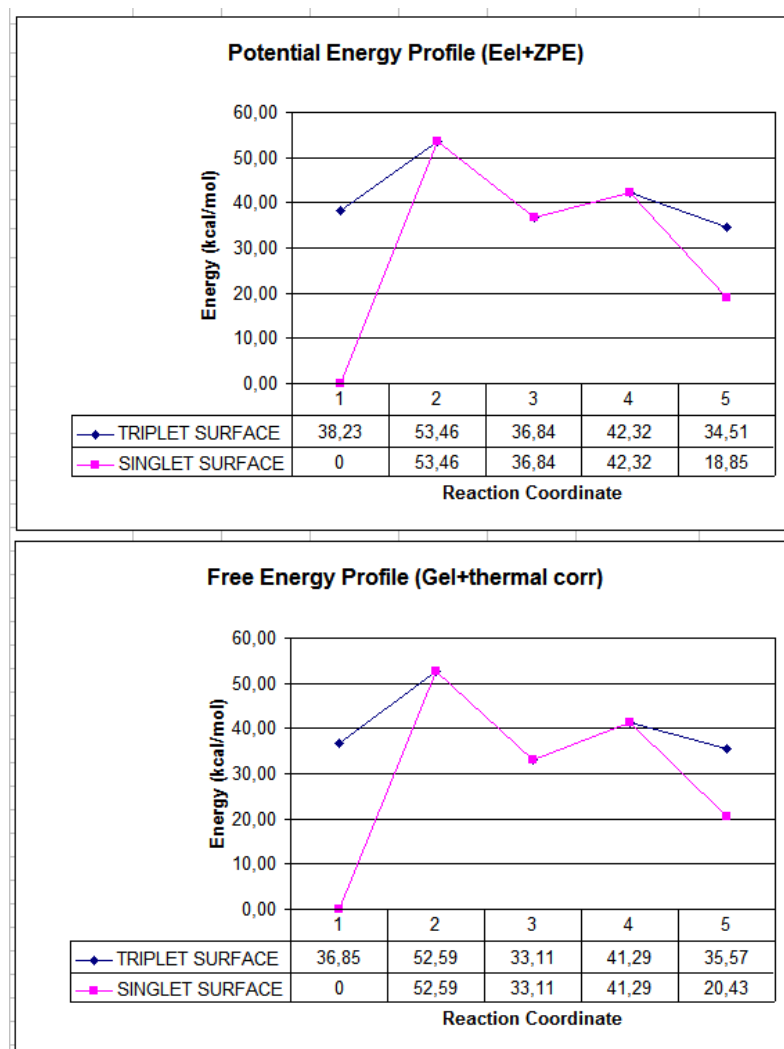


Figure 4. 9: Potential and Gibb’s free energy profiles of singlet and triplet surfaces of ‘Two-water’ system. (See text for definition). Obtained at B3LYP/6-31g* level. The energy values are adjusted with respect to the energy of Structure 1 at singlet surface.

In Figure 4.9, it may be observed that passing from singlet surface to triplet surface decreases the energy barrier for the first step by about 37 kcal/mol as in the case of ‘Without-water’ system. At triplet surface, activation free energies of the first and the second steps are 15.74 kcal/mol and 8.18 kcal/mol, respectively. Based on this energy barrier values, the first step is the rate-determining step. This is contradicting with the results of ‘Without-water’ and ‘Single-water’ systems. At the end of Step 2, switching back to singlet surface is more favorable as triplet surface optimized energy is ca. 15 kcal/mol. This is consistent with what was observed for ‘Without-water’ system. When the energy profiles of ‘Without-water’,

'Single-water' and 'Two-water' systems are compared, it is seen that, 'Two-water' system has higher energies for all structures, except the Structure 5. In this frame, the presence of two water molecules seems to be destabilizing the reaction. However, we should look at the effect of the molecular oxygen on the reaction in order to have a safe understanding of the effect of water molecules.

4.2.3 'O₂-Two-Water' and 'MonoMethyl' Systems

4.2.3.1 'O₂-2H₂O' System

As the final step of only-QM calculations, a system was prepared that contains dimethyl lysine, flavin ring, two crystal water molecules and single molecular oxygen, i.e. O₂. This system will be referred to as 'O₂-2H₂O' system in this thesis from this point on.

For the start-up, single molecular oxygen is optimized at singlet and triplet surfaces in order to check the *Stability* calculation results as explained in Chapter 3. The optimized geometry at the triplet surface has a lower energy value by 41.77 kcal/mol. This shows that there are two unpaired electrons on optimized O₂ structure. This will be considered while calculating the spin multiplicity of whole system.

As the next step, the orientation of the molecular oxygen with respect to other molecules was tried to be determined. For this purpose, PM3 orientation and docking procedures were carried out once more.

For the PM3 orientation process, 21 different trials were prepared. In these trials, flavin and lysine were kept fixed and the two water molecules were put near to the positions shown in the crystal structure (PDB ID: 2H94). The position of the molecular oxygen was changed systematically and arbitrary geometries were prepared. The best orientation based on the energy values were used as the input structure for further optimization at B3LYP/6-31g* level on singlet, triplet and quintet surfaces. Among three surfaces, being on the triplet energy surface stabilizes the system at most. Triplet surface has a lower potential energy value by 38 kcal/mol and 45 kcal/mol compared to other surfaces, i.e. singlet and triplet, respectively, (Table 4.5). This is consistent with expectations. As the system contains molecular oxygen, there are two unpaired electrons in the system and they require triplet multiplicity to be stable.

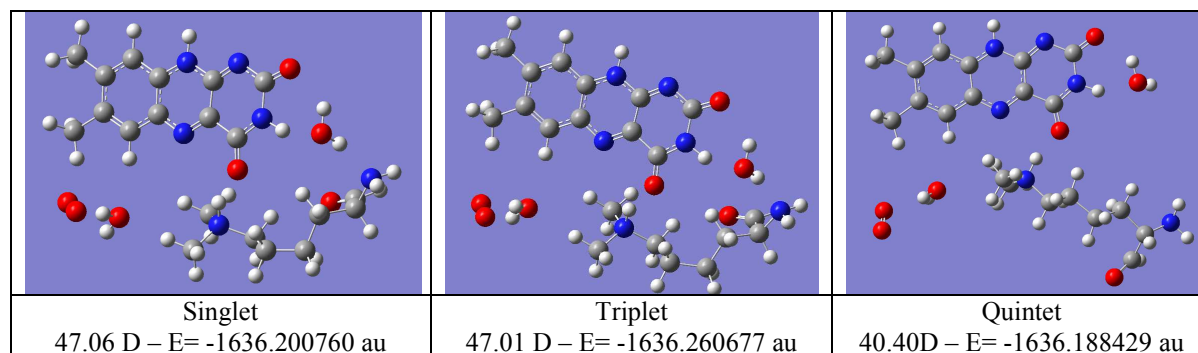


Table 4. 5: Optimized geometries of Step 1 reactants of ‘O₂-2H₂O’ system obtained on singlet, triplet and quintet surfaces. Obtained at B3LYP/6-31g* level. See text for system definition. Energy and dipole moments of the structures are also given.

After the position of molecular oxygen was determined, whilst molecular oxygen was kept fixed, positions of the water molecules were changed and the energy change was observed. The best energy results were obtained when the two water molecules are near to the positions indicated in the crystal structure.

On contrary, the position of molecular oxygen was also determined via docking procedure with *AutoDock*. In this procedure, first the lysine was docked on the macromolecule that consists of FAD, neighboring enzyme residues and two water molecules using *LGA* method with a binding energy of -2.12 kcal/mol. Afterwards, molecular oxygen is docked on the macromolecule obtained in the previous step. The molecular oxygen was docked to the position shown in Figure 4.10 using *LGA* method with a binding energy of -0.44 kcal/mol, which corresponds to not a very strong binding. Nevertheless, as it is seen from Figure 4.10, the molecular oxygen is stabilized through hydrogen-bonds it makes with flavin and crystal water. In fact, loose binding of O₂ is required, since the molecular oxygen needs to be mobile, while it takes the hydrogen atoms of the FADH₂ during the reaction.

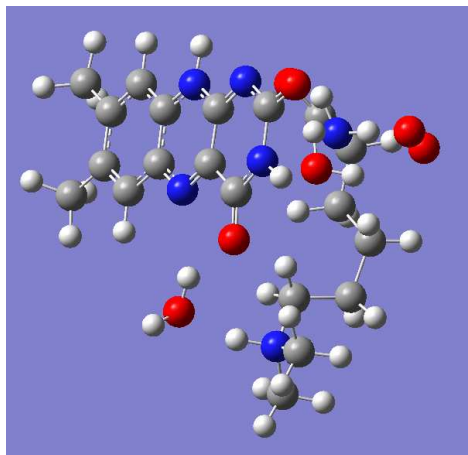


Figure 4. 11: B3LYP/6-31g* level triplet surface optimized geometry of Step 1 reactants docked by *AutoDock*. E= -1636.259306 and 12.73D

For optimizing the rest of the stationary points shown in the reaction mechanism (Figure 4.1), PM3 oriented structure was modified appropriately and used in preparing input structures.

Before giving the energy profiles for ‘O₂-2H₂O’ system, some observations for the stationary points will be given. First, the optimized structure of Step 1 products (i.e. Structure 3 in Figure 4.1) was observed to have the dimethyl lysine beneath the flavin ring, on contrary to other prepared systems, i.e. ‘Without-water’, ‘R=CH₃’, ‘Single-water’ and ‘Two-water’. In those systems, lysine tends to shift to the top of flavin ring in order to align the methyl carbon atom of lysine and the nitrogen atom of the flavin for the hydride transfer occurring in the second step. Moreover, when two crystal water molecules are present in the system, the water in the right-hand side of the flavin accompanies the lysine and shifts to the top of the flavin along with lysine.

On contrary, in ‘O₂-2H₂O’ system, during optimization the path of lysine towards to the top of the flavin is blocked by molecular oxygen. Therefore, it stays beneath the flavin ring. (Table 4.6-a). We wondered if pulling the lysine to the top of the flavin ring would reduce the energy. Therefore, we pulled only the lysine to the top in one trial (Table 4.6-b), while in another we pulled the right-hand-side water next to lysine (Table 4.6-c) and optimized these trials at B3LYP/6-31g* level. As the result, pulling the lysine to the top of the flavin (Table 4.6-b) stabilizes the system by 4.63 kcal/mol. Moreover, the system is further stabilized by the shift of the water molecule next to the lysine (Table 4.6-c) by 4.57kcal/mol. This corresponds to a total of 9.2 kcal/mol free energy stabilization at B3LYP/6-31g* level,

while 7.22 kcal/mol at MP2-level. On the light of these observations, it may be inferred that at the end of Step 1 lysine is pushed to the top of the flavin ring and one water molecule (i.e. the one found on the right hand-side of flavin ring) accompanies lysine along its shift.

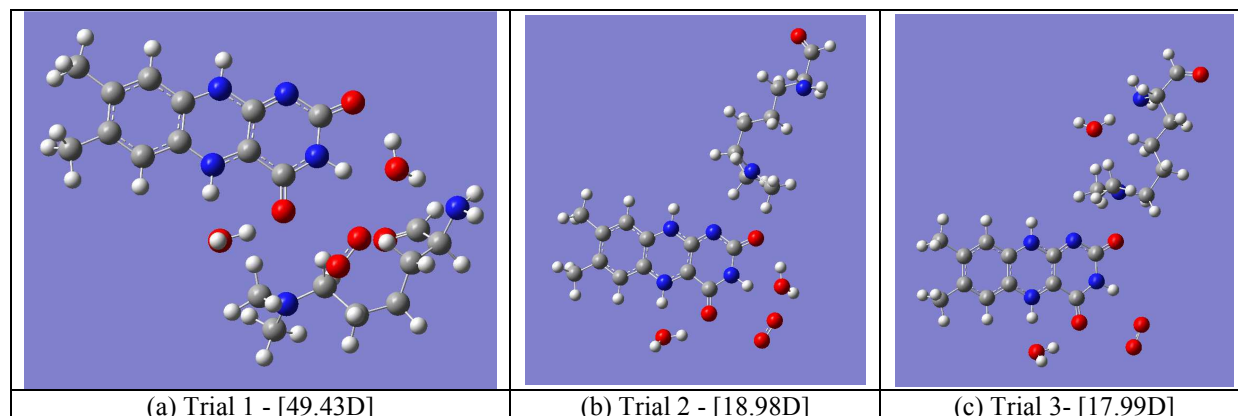


Table 4. 6: Optimized structures for Step 1 products of 'O₂-2H₂O' system. Obtained at B3LYP/6-31g* level.

On the other hand, after the molecular oxygen is converted to hydrogen peroxide, i.e. H₂O₂ (Structure 6), a water molecule is needed for hydration of aminium cation in Step 3. We hypothesized that the right-hand-side water molecule is more likely to be consumed during this process rather than the water molecule found beneath the flavin ring. That is because the right-hand-side water molecule is getting closer to the lysine compared to the other water molecule, when the lysine shifts to the top of the flavin at the end of Step 1; therefore it is more likely to hydrate the aminium cation. To test out the hypothesis, two different trials were prepared for the Step 4 products (i.e. Structure 10). In one trial, the single water molecule is placed beneath the flavin; while in the other, the single water molecule is placed in the right-hand side of the flavin (to the positions shown in the crystal structure). As the result, when the water is placed beneath the flavin ring, system has a lower free energy at singlet surface by 6.66 kcal/mol. (Optimized geometries may be found in Table 4.7). That may be accepted as a proof that the water on the right-hand side is consumed during the reaction.

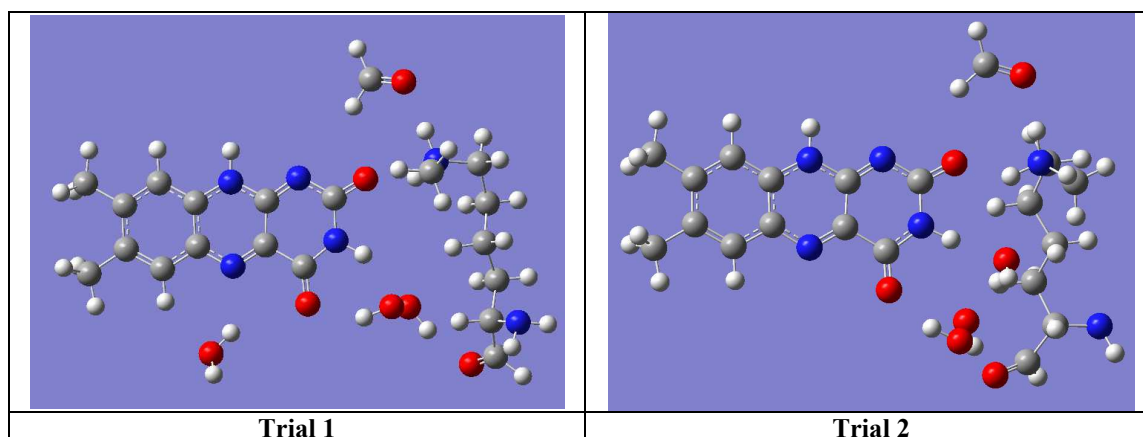


Table 4. 7: Optimized geometries for the two trials for determining the favorable position of single water at the end of Step 4 (obtained at singlet surface and B3LYP/6-31g*). In Trial 1, water molecule is beneath the flavin ring, in Trial 2 it is at the right-hand side of flavin ring.

The optimized geometries of the TS structures for Step 1 and 2 obtained at B3LYP/6-31g* level are given in Table 4.8. The TS structures could be optimized only at the quintet surface, while the trials with other energy surfaces, e.g. singlet or triplet, were not successful. This was expected, since TS structures contain radicals and molecular oxygen, as was the case in previous systems. The negative frequency values of TS1 and TS2 are -885.13 cm^{-1} and -914.82 cm^{-1} , respectively. When compared to the negative frequency values obtained with ‘Without-water’ system (928.74 cm^{-1} and -1024.24 cm^{-1} , respectively), the negative frequency values ‘O2-2H2O’ system follow the same trend for TS1 and TS2, but a little higher in value. This may be caused by the effect of the presence of two water molecules and molecular oxygen. In addition, the geometries correspond to the imaginary movement of hydrogen atoms, which hang between the nitrogen and carbon atoms of lysine and flavin.

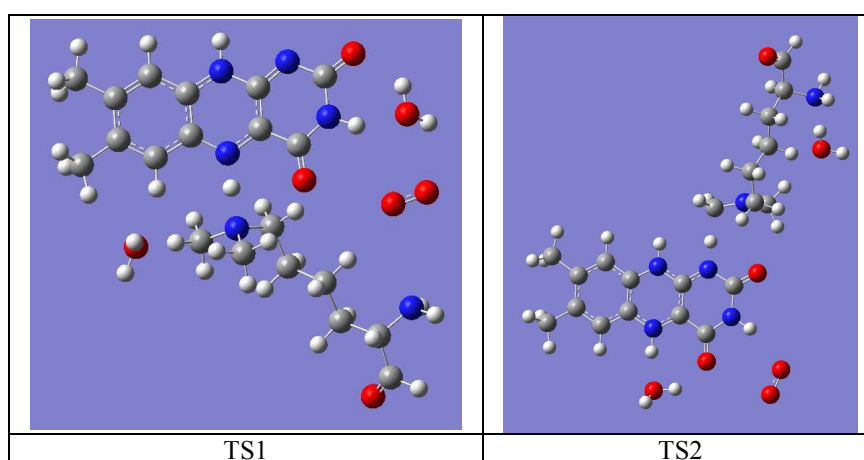


Table 4. 8: Optimized geometries of transition state structures of Step 1 and 2 of ‘O2-2H2O’ (see text for system definition). They are obtained with the multiplicity value of 5 (Quintet Surface) and at B3LYP/6-31g*.

For obtaining the energy profiles, structures were optimized with singlet, triplet and/or quintet multiplicity. For clearing the ambiguity, the multiplicity values are given for each structure as follows. Structure 1: singlet, triplet and quintet; Structure 2: quintet; Structure 3: quintet; Structure 4: quintet; Structure 5: triplet and quintet, Structure 6: singlet, Structure 8: singlet; Structure 10: singlet, triplet and quintet. In the energy profiles three energy surfaces are presented. If a structure was optimized with only one multiplicity value, energy values of the other surfaces are shown to be the same as the energy value of the only available surface (although this is not the case in real). For example, Structure 2, 3, 4 were optimized only with quintet multiplicity. Therefore, only one energy value is available for each structure. Besides, TS structures of Step 3 and 4 (i.e. Structure 7 and 9) were not optimized, as they are not important in terms of enzyme activity. That means, since the rate-determining steps are step 1 or 2; the enzyme is involved only in these steps and thus analyzing the TS structure of Step 3 and 4 is not included in the scope of this thesis. They may also be optimized as future work.

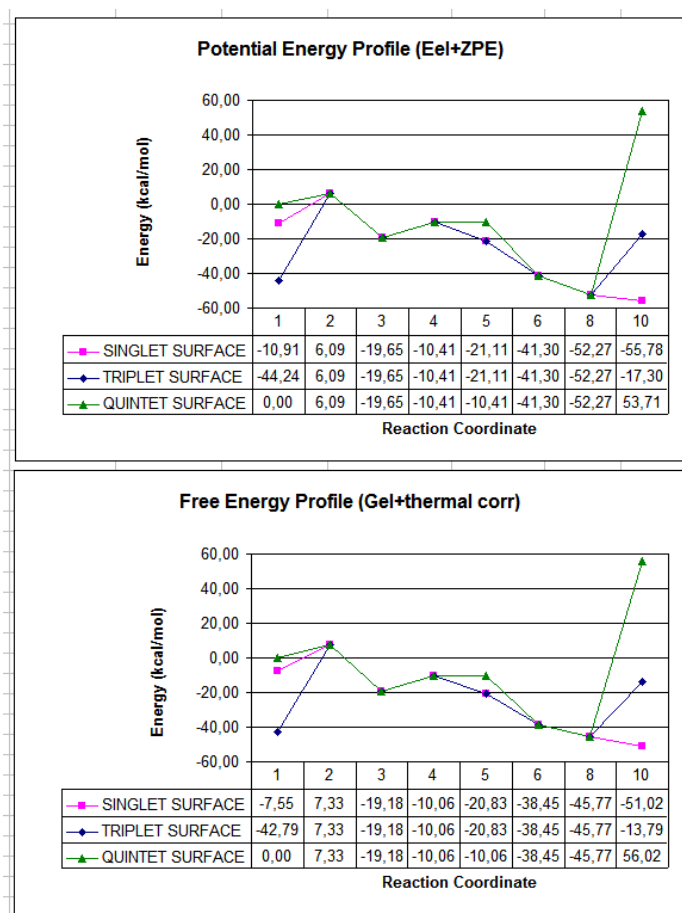


Figure 4. 12: Free energy and potential energy profiles of singlet, triplet and quintet surfaces of ‘O2-2H2O’ system (see text for definition) obtained at B3LYP/6-31g* level.

If the energy profiles of 'O₂-2H₂O' system (see Figure 4.12) are analyzed, it is seen that only Structure 1, 5 and 10 could be optimized at different energy surfaces. Therefore, only these structures are compared. Structure 1 has the lowest energy when optimized at triplet surface compared to ones obtained at singlet and quintet surfaces. This is consistent with our expectations as the system contains molecular oxygen, which is shown to have lower energy at triplet surface.

Structure 2, 3 and 4 were optimized only at quintet surface. These structures could not be obtained at triplet or singlet surface, as they involve radicals that have unpaired electrons. In agreement with this observation, Structure 5 could be optimized at triplet and quintet surfaces, but not on singlet surface, because of the existence of molecular oxygen. The geometry optimized at triplet surface has a lower energy by ca. 21 kcal/mol (ca. 50kcal/mol at MP2/6-31g* level). This may be inferred as that at the end of Step 2, the reaction jumps back to triplet surface from quintet surface in order to decrease energy.

Structure 6 corresponds to the case when the molecular oxygen is oxidized by taking two protons of FAD. Since the valance orbitals of molecular orbitals are filled with electrons coming from protons, there remain no unpaired electrons anymore, so the multiplicity must be equal to 1. Even if it is not shown in Figure 4.12, optimized geometry of Structure 6 has a higher energy by ca. 27 kcal/mol at triplet surface compared to singlet surface. This shows geometry of Structure 6 prefers to be at singlet surface, as our reaction mechanism proposal offers.

In addition, when the FADH₂ is reduced by molecular oxygen, the energy of the system decreases by about 18 kcal/mol. This may be interpreted as that reduction of FADH₂ back to its original form, i.e. FAD, is preferably occurring before the hydration of aminium cation is carried out (i.e. Step 3 in Figure 4.1).

The last two steps of the reaction are proposed in the chemical mechanism (i.e. Figure 4.1) to be occurring spontaneously, i.e. without requiring enzyme activity, at the singlet surface. The energy barriers of these steps were not calculated, but they are considered to be very low compared to the energy barriers of the first and second steps. The products of the last step were optimized at singlet, triplet and quintet surfaces and the lowest energy was obtained at the singlet surface. This information can be interpreted as the reaction system should not remain at the quintet surface until the end of the reaction in order to obtain the

lowest possible energy. Otherwise (i.e. if the reaction does not switch back to singlet surface), it will create the destabilization of ca. 108 kcal/mol. As observed from the energy profiles, with these multiplicity value choices (i.e. Structure 1 is singlet, triplet or quintet, and Structure 10 is singlet), the reaction is exothermic, thus is energetically favorable.

On the other hand, the reaction prefers to be at the triplet surface at the beginning of the Step 1 in terms of energy. However, the reaction may be dragged to quintet surface in the presence of the enzyme in order to decrease the energy barrier of the first step. The activation free energy of the first step is reduced from ca. 50 kcal/mol to 7.33 kcal/mol when the reaction jumps from triplet surface to quintet surface. In this case, the second step becomes the rate-determining step by having an energy barrier of 9.12 kcal/mol. When compared to 'Without-water' system, energy barrier of Step 1 and 2 are decreased in 'O₂-2H₂O' system by 6.48 kcal/mol and 3.94 kcal/mol, respectively. This proves that the presence of molecular oxygen and two crystal water molecules is required for the reaction, as they stabilize the reaction system and enhances the reaction rate.

The energy barrier of the rate-determining step corresponds to the reaction rate, i.e. $7.65 \times 10^7 \text{ min}^{-1}$. The reaction rate is very high (in the order of 10^7) compared to the experimental result. This was expected, since the DFT methods were shown to be creating lower energy barriers compared to experimental values. (Please refer to Chapter 3). Therefore, we calculated the energy profile using single point MP2/6-31g* energy calculation results on the optimized geometries that were previously obtained via B3LYP method.

In Figure 4.13, the energy profile obtained at MP2/6-31g* level is given. While preparing the energy profile following multiplicity values were used: Structure 1: singlet, triplet and quintet; Structure 2: quintet; Structure 3: quintet; Structure 4: quintet; Structure 5: triplet and quintet, Structure 6: singlet, Structure 8: singlet; Structure 10: singlet.

Structure 1 has the lowest energy at triplet surface, which is consistent with the previous calculations. The first energy barrier is very small –almost non-existent- (0.36 kcal/mol) at the quintet surface (i.e. in the presence of enzyme) compared to triplet surface energy barrier (ca. 89 kcal/mol). In contrast, the second step is observed to be the rate-determining step by having an energy barrier of 12.91 kcal/mol. That is consistent with previous calculations, i.e. B3LYP calculations. This energy barrier corresponds to the reaction rate of $1.27 \times 10^5 \text{ min}^{-1}$, which is more realistic, i.e. closer to the experimental value,

compared to one obtained at B3LYP/6-31g* level. However, this value still corresponds to a high reaction rate. Moreover, Structure 10 has lower energy compared to all energy values obtained at singlet, triplet or quintet surfaces for Structure 1. That means the reaction based on the proposed chemical mechanism is exothermic, so energetically favorable and the mechanism is likely to be valid for describing the demethylation process.

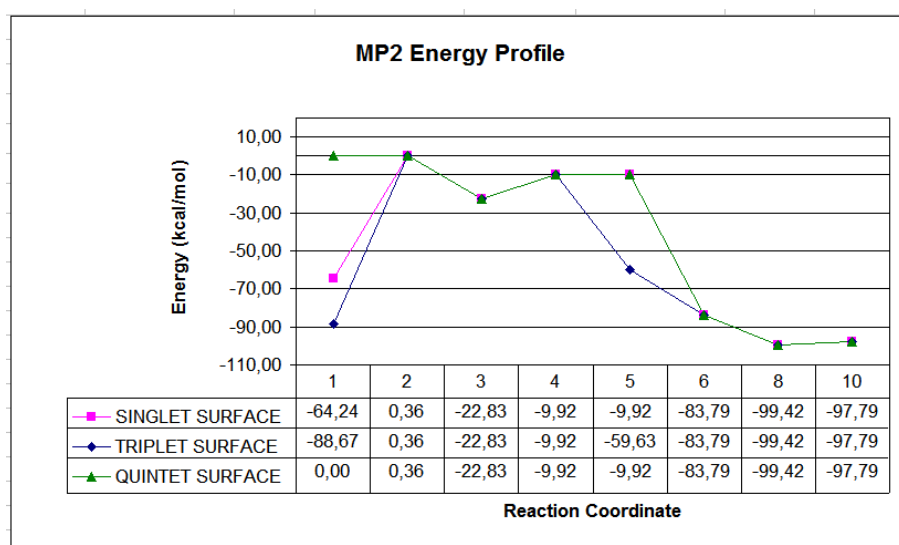


Figure 4. 13: Energy profile for ‘O2-2H2O’ system obtained at MP2/6-31g* level. Three energy surfaces are present: Singlet, triplet and quintet.

4.2.3.1 ‘MonoMet’ System

As the next step, demethylation of monomethyl lysine was investigated in order to compare it with the system prepared for dimethyl lysine demethylation reaction. For this purpose, the dimethyl lysine system, i.e. ‘O2-2H2O’ system, was taken as basis and some minor changes were done. Namely, one methyl group of lysine was changed into hydrogen and the rest of the system, i.e. flavin, molecular oxygen and two crystal water molecules, was kept as is. All the stationary points were modeled based on the same chemical mechanism offered for dimethyl lysine demethylation (Figure 4.1) only the dimethyl lysine derivatives were replaced with monomethyl lysine derivatives. This system prepared for monomethyl demethylation will be referred to as ‘MonoMet’ system from this point on in this thesis.

The optimized geometries of the transition state structures for ‘MonoMet’ system obtained at quintet surface and B3LYP/6-31g* level are given in Table 4.9.

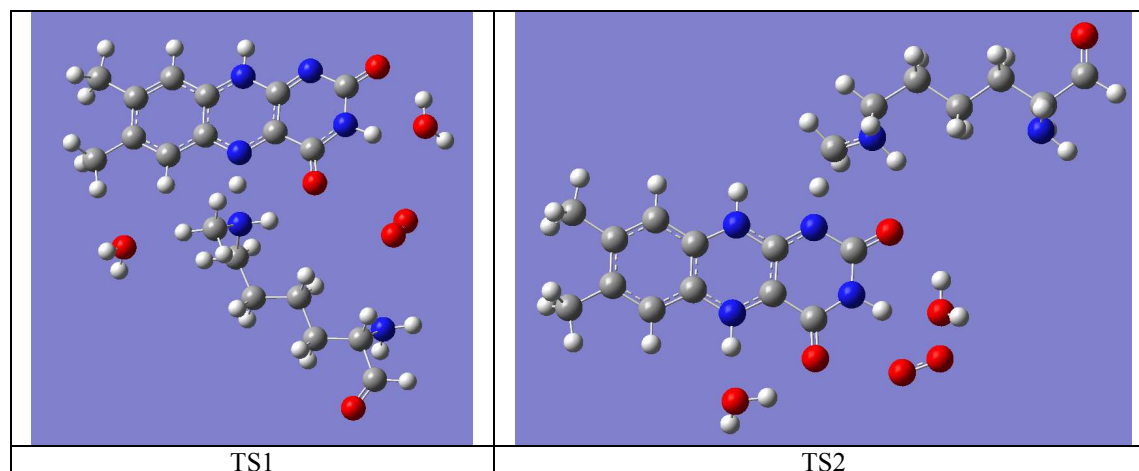


Table 4. 9: Optimized geometries of transition state structures of Step 1 and 2 of ‘O2-2H2O’ (see text for system definition). They are obtained with the multiplicity value of 5 (Quintet Surface) and at B3LYP/6-31g*.

Negative frequency values for optimized geometries of TS1 and TS2 for ‘MonoMet’ system, which were obtained at quintet energy surface, are -528.13 cm^{-1} and -1038.89 cm^{-1} , respectively (Table 4.10). Comparing the values with the ones obtained for ‘O2-2H2O’ system, it is seen that the negative frequencies obtained for TS1 and TS2 of ‘MonoMet’ are very low and high, respectively, compared to the ones obtained for ‘O2-2H2O’. Different negative frequencies correspond to different imaginary vibrations. Higher frequency vibrations make more rapid movement, but with a smaller range, which means total displacement of all atoms found in the system is smaller during the vibration. In this light, involving higher negative frequency (-528.63 cm^{-1}), the vibration of TS1 of ‘MonoMet’ system is more rapid, which may be correlated with the lower energy barrier (i.e. higher reaction rate) compared to ‘O2-2H2O’ system (3.67 kcal/mol vs. 7.33 kcal/mol) that can be observed in the energy profiles (Figure 4.12). Besides, the lower negative frequency value of TS2 of ‘MonoMet’ system compared to ‘O2-2H2O’ may be correlated with the higher energy barrier for Step 2 of ‘MonoMet’ system.

System name	“O2-2H2O”	“MonoMet”
Transition State of Step 1	-885.13 cm^{-1}	-528.63 cm^{-1}
Transition State of Step 2	-914.82 cm^{-1}	-1038.89 cm^{-1}

Table 4. 10: Negative frequency values for the optimized TS1 and TS2 geometries of ‘O2-2H2O’ and ‘MonoMet’ systems. (See text for system definitions). (Optimization at B3LYP/6-31g* level)

Although TS structures of both systems involve basically the same hydrogen transfer motion, the motion is affected by the neighboring groups that are bound to nitrogen atom of the methylated lysine. In details, the nitrogen atom on the monomethyl lysine has one methyl group and hydrogen. In contrast, dimethyl lysine has two methyl groups bound to the nitrogen atom, which makes it unfeasible for the corresponding atoms on the flavin and lysine to get closer in TS1 and TS2. This can be observed from the related bond lengths and angles given in Table 4.11. As clearly seen from the table, in optimized TS1 and TS2 geometries of dimethyl lysine N-N and N-C bond lengths are ca. 0.03-0.04 Å longer compared to monomethyl lysine.

System Name		'O2-2H2O'	'MonoMet'
TS1	N-N distance	2.66040 Å	2.6394 Å
	N-H-N angle	173.333°	173.786°
TS2	N-C distance	2.74009 Å	2.70010 Å
	N-H-C angle	178.396°	166.746°

Table 4. 11: Some key structural properties for TS1 and TS2 structures of 'Without-water', 'O2-H2O', 'MonoMet' systems (see text for system descriptions).

The potential and free energy profiles for 'MonoMet' system were prepared using the following multiplicity values for each stationary point shown in the chemical mechanism. Structure 1: singlet, triplet and quintet; Structure 2: quintet; Structure 3: quintet; Structure 4: quintet; Structure 5: triplet and quintet; Structure 6: singlet and triplet; Structure 8: singlet; Structure 10: singlet. Energy profiles that were obtained at B3LYP/6-31g* level are given in Figure 4.14.

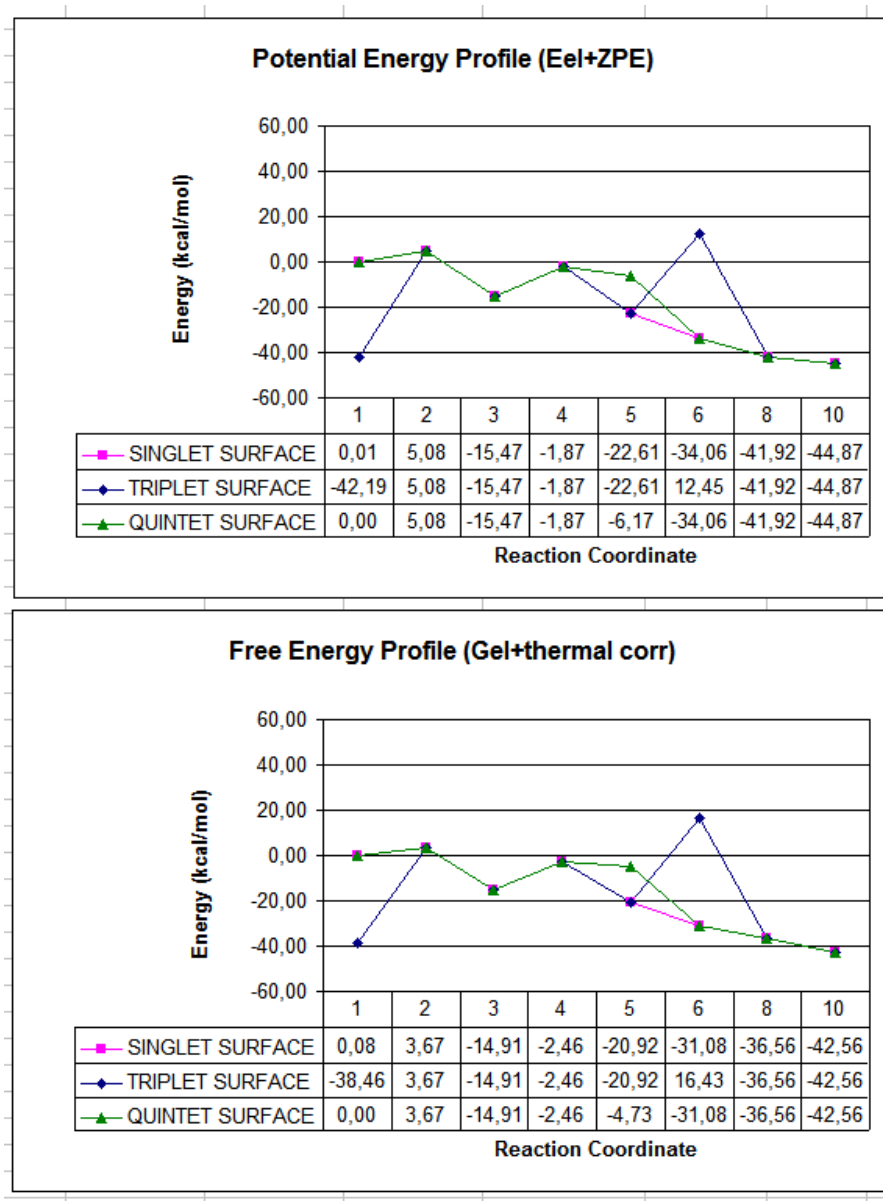


Figure 4. 14: Potential and Gibb’s free energy profiles of singlet, triplet and quintet surfaces of ‘MonoMet’ system (see text for system definition) obtained at B3LYP/6-31g* level.

The following observations can be deduced from the energy profiles. The most important observation to be made is that the mono methyl lysine demethylation reaction is exothermic as in the case of dimethyl lysine demethylation. Therefore, the reaction is energetically favorable and the chemical reaction mechanism is likely to be valid for the demethylation of monomethyl lysine.

Energy of the Step 1 reactants, i.e. Structure 1, are the lowest, when it is optimized at triplet surface compared to singlet and quintet surfaces, as observed in the previous systems,

i.e. ‘O₂-2H₂O’ and other systems. The free energy barrier of the first step is ca. 42 kcal/mol when the Step 1 reactants are at the triplet energy surface and it decreases to 3.67 kcal/mol, when both Structure 1 and 2 are at quintet surfaces. In this case, the second step is the rate-determining step by having an energy barrier of 12.45 kcal/mol. This corresponds to the reaction rate of 2.76×10^5 , which is 10^5 -fold higher than the experimental value ($3.40 \pm 0.1 \text{ min}^{-1}$). This is the same case observed in B3LYP/6-31g* level energy profile of dimethyl lysine system.

Like in the profiles of ‘O₂-2H₂O’ system, reaction starts at quintet surface and jump back to triplet surface at the end of Step 2 (based on the energy difference of ca. 16 kcal/mol). Then reaction jumps to singlet surface in the intermediate step at which molecular oxygen is reduced to peroxide (based on the energy difference of ca. 47kcal/mol). The Step 3 and 4 occur at singlet surface.

For obtaining a more accurate energy profile, MP2 level single point calculations were done. The same multiplicity value selections were applied to each stationary point, as were in the B3LYP optimizations.

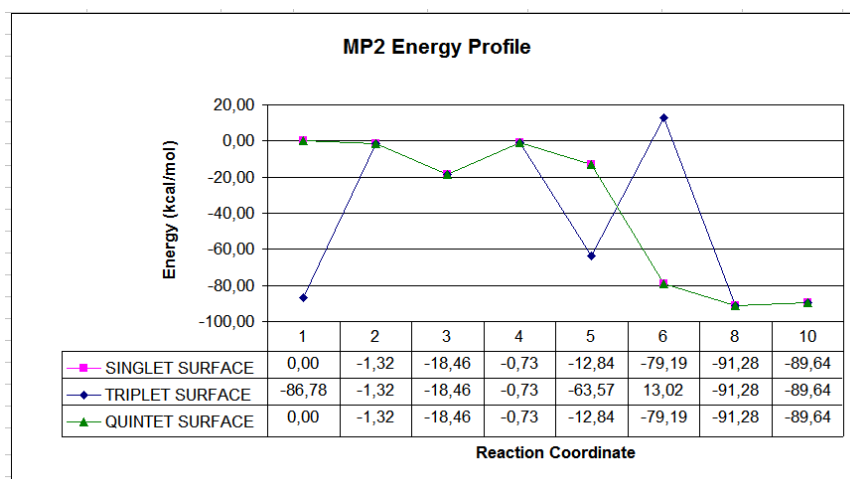


Figure 4. 15: Energy profile for ‘MonoMet’ obtained at MP2/6-31g* level. (See text for system definition).

In Figure 4.15, the same trend is observed as in the B3LYP/6-31g* level energy profile. The energy difference of triplet surface with singlet and quintet surfaces for Structure 1 is more significant (ca. 87 kcal/mol) compared to B3LYP-level values. Moreover, energy barrier for Step 1 is indicated as 85.46 kcal/mol at triplet surface and 1.32 kcal/mol at quintet surface. In contrast, energy barrier of Step 2 is 17.73 kcal/mol, so that the second step is the rate-determining step. The calculated reaction rate is 37.02 min^{-1} and very close to the

experimental value ($3.40 \pm 0.1 \text{ min}^{-1}$) compared to B3LYP-level result for monomethyl lysine. This shows that proposed chemical mechanism is very likely to be suitable for demethylation of monomethyl lysine.

Comparing the reaction rates of monomethyl and dimethyl lysine systems ('O2-2H2O' and 'MonoMet'), dimethyl lysine system has higher reaction rate at B3LYP- and MP2-level compared to monomethyl lysine system. This is consistent with the experimental reaction rates, which allege that LSD1 is more active in demethylation of dimethyl lysine compared to monomethyl lysine.

4.3 QM/MM Simulation System and Calculation Results

After finishing the analysis with only-QM systems, we carried out the hybrid QM/MM system calculations, preparation of which is explained in Chapter 3, in order to understand the effect of the remaining part of the FAD and neighboring enzyme residues on the demethylation reaction of dimethyl lysine.

In all of the representations of the QM/MM systems given in the figures and tables in this subchapter, QM region and MM region atoms are depicted with ball-and-stick and wireframe (lines) representations, respectively.

4.3.1 'OnlyFAD' System

As the start-up for the QM/MM calculations, a system was prepared by using the AutoDock results, as described in the subchapter related to 'O2-2H2O', carried out for docking the O₂ on the macromolecule that contains the FAD molecule, two water molecules and the neighboring enzyme residues. In contrast to the preparation of 'O2-2H2O' system, not only the flavin ring of the FAD molecule but whole of the molecule is taken into consideration. In that sense, this system is different from 'O2-2H2O' system by containing rest of the FAD molecule.

For clearing any possible ambiguity, flavin ring, dimethyl lysine, two water molecules and molecular oxygen are placed in the QM part, while the rest of FAD molecule is placed in the MM part. This system will be referred to as 'OnlyFAD' system from this point on in this thesis. The charge of the MM part is fixed to zero, i.e. neutral.

On contrary to previously introduced systems, no further orientation work was done for the Structure 1, since the structure previously obtained from docking procedure with *AutoDock* (Figure 4.17) is used as Structure 1 of the chemical mechanism.

Before giving the energy profiles related to ‘OnlyFAD’ system, some observations related to geometry optimizations of the stationary points are given.

Similar to what was done for ‘O2-2H2O’ and ‘MonoMet’ systems, two trial input geometries were prepared for Step 1 products, i.e. Structure 3 (Table 4.12). In Trial 1, lysine is beneath the flavin ring; while in Trial 2, lysine is placed to top of the flavin ring. This was done for determining if the lysine prefers to be dragged to the top of the flavin ring, where its methyl carbon atom is close to nitrogen atom of the flavin ring, which is required for the second step to proceed. When the energies of the optimized geometries of two trials are compared, it is seen that it is energetically favorable (by 26.60 kcal/mol) for lysine to be located close to the top of the flavin ring. This can be interpreted as lysine is migrating upwards at the end of Step 1 for preparation of the next step. Besides, a water molecule accompanies the lysine during the migration, which is used in hydration of aminium cation during Step 3. This is consistent with what is observed in previously-analyzed only-QM systems.

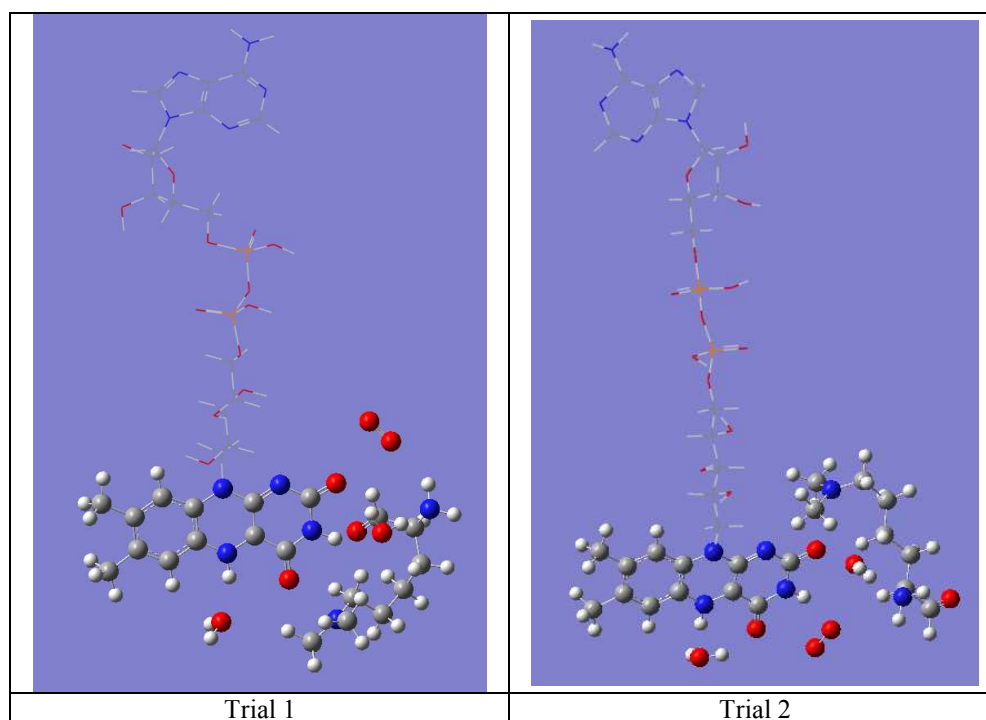


Table 4. 12: Optimized geometries for Trial 1 and 2 prepared for Step 1 products of ‘OnlyFAD’ system. (See text for system definition). Optimized at quintet surface and B3LYP/6-31g* level.

As already explained previously, molecular oxygen is reduced to hydrogen peroxide just after Step 2. Eight different inputs were prepared for determining the position of the molecular oxygen and hydrogen peroxide at which the most stable geometry in terms of energy is obtained. In these trials, molecular oxygen is placed with different orientations near to two hydrogen atoms of FADH₂, which will be transferred to molecular oxygen, and all structures optimized at B3LYP/6-31g*:UFF level. Successively, the optimized geometries are modified so that FADH₂ and O₂ are converted in to FAD and H₂O₂, and these geometries are used as input structures in geometry optimization at the same level. The least possible energy (the most stable structure) is obtained when the molecular oxygen is placed in the position determined by the docking procedure with *AutoDock*. Therefore, the energy profiles were prepared with the stationary points having molecular oxygen or hydrogen peroxide near to the positions determined with *AutoDock*.

Proper geometry optimization for transition state structures of Step 1 and Step 2 could not be achieved even though numerous input geometries, which were prepared with different approaches, were tried. In almost all of the geometry optimization trials, optimization converged to products of the corresponding step. In very few trials, optimization converged to reactants, as well.

Besides, on contrary to TS1, we achieved to obtain an optimized geometry, which resembles the desired TS structure, for TS2. Corresponding geometry is given in Table 4.13. Nevertheless, this structure yielded two imaginary vibrations (rather than single), none of which correspond to the desired hydrogen transfer movement between nitrogen atom of flavin and carbon atom of lysine's methyl group. Besides, this geometry yielded potential and Gibb's free energy values lower than the Step 2 reactants, i.e. Structure 3. This contradicts with the Transition State theory that alleges a TS structure must have higher energy compared to the reactants and products it is connecting. Therefore, this geometry cannot be the desired TS2 structure geometry.

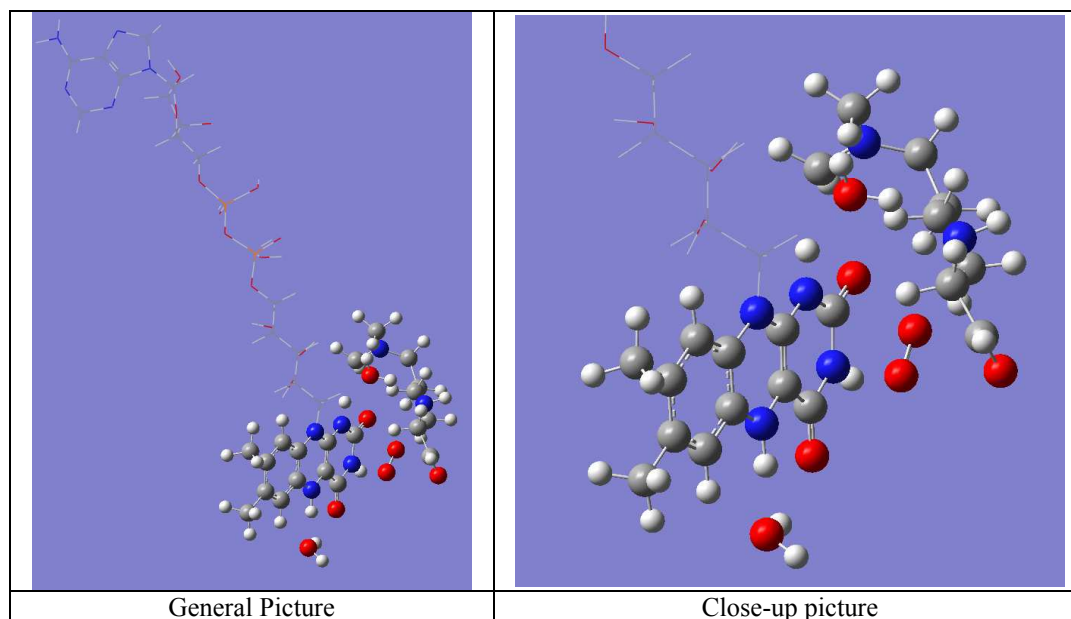


Table 4. 13: Optimized geometry of TS2-like structure for 'OnlyFAD' system obtained with multiplicity value of 5 and at B3LYP/6-31g*:UFF level (see text for system definition).

Optimization with higher basis set (i.e. 6-31+g*) using the aforementioned optimized geometry obtained at B3LYP/6-31g* did not succeed as the optimization converged again to products.

The trials for optimizing the TS structures for Step 1 and 2 did not succeed, because of the possible reasons explained in the following paragraphs.

Optimizations, commonly, converged to products of the corresponding step. This can be inferred as the TS structures have energies very close to the energy of the reactants, which lead to very low energy barriers, therefore during optimizations these energy barriers are easily overcome and the algorithm converges to the products, which are energetically more favorable compared to reactants. (The energy values of reactants and products can be referred to in the energy profile given in Figure 4.16).

This hypothesis is consistent with the Hammond's postulate, which proposes that TS structure of an exothermic process resembles the reactants in terms of geometry and energy (and other thermodynamics properties, as well). In this frame, as TS resembles the reactants, their energies are very similar, forming a plateau on the energy surface. The products are bound to this plateau with a sharp slope, so that during the optimization, algorithm falls frequently to the products' pit on the energy surface. Besides, in very few cases, algorithm converges to the reactant-like structures based on the input structures, which correspond to the

points on the energy surface so much close to the reactants' local minimum that the algorithm is trapped to the local minimum instead of finding the so-called global minimum corresponding to products.

Another reason for the failure in geometry optimization of any TS structure for 'OnlyFAD' system may be the low performance of ONIOM method and/or method duos (i.e. B3LYP-UFF combination) in simulating the radical systems. Some supporting findings for this hypothesis are the previous studies using DFT methods, which suggest that DFT methods are insufficient to some extent in simulating the electron interactions and some physical key interactions, as explained in Chapter 3. As the radicals are sensitive to electron interactions more than the stable molecules, DFT methods may be impotent to simulate the interaction between the radicals. However, as the B3LYP:UFF combination is successful in simulating Structure 3, which also includes radicals but a minimum point rather than a TS point. This means, ONIOM method may be insufficient for modeling the TS structures, which include radicals.

On contrary, B3LYP:UFF combination and ONIOM method were used in some previous studies [1, 117, 139-151] and referred to as reliable for simulating organic systems. Therefore, this reason is less likely compared to the first reason.

Gibb's free energy and potential energy profiles are presented in Figure 4.16. In the energy profile TS points, i.e. Structure 2 and 4, are missing as explained previously. While preparing the energy profiles, the following multiplicity values were used. Structure 1: quintet; Structure 3: quintet; Structure 5: triplet; Structure 6: singlet; Structure 8: singlet; Structure 10: singlet. All of the structures were optimized only at a single surface at which the optimized geometry has the lowest energy. These surfaces were determined based on the previous observations made in the energy profiles of only-QM systems.

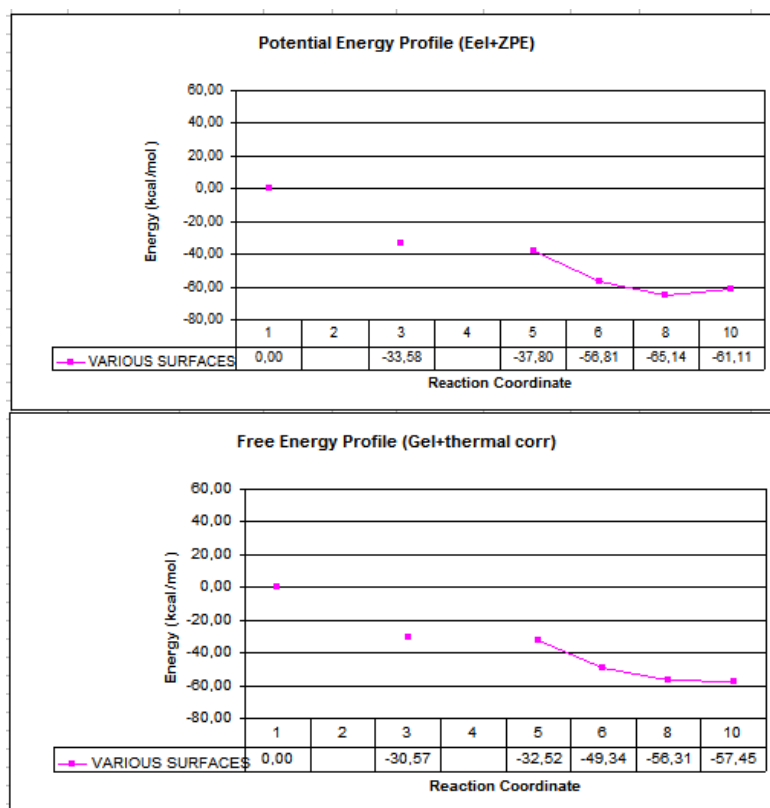


Figure 4. 16: Potential and Gibb's Energy profiles for 'OnlyFAD' system (See text for system definitions).

In the energy profile, it is seen that the exothermic character of the reaction does not change with the addition of the rest of the FAD molecule. As the TS points are missing in the profile, the energy barriers and the reaction rate of 'OnlyFAD' system and the previous systems cannot be compared. Nonetheless, the energies of **all** available points on the profiles (relative to Structure 1) of 'OnlyFAD' system are lower than the ones corresponding to 'O2-2H2O' system. This may be interpreted as the tail of FAD (the FAD molecule without the flavin ring) stabilizes all of the stationary points on the chemical mechanism.

The tail part of FAD was previously assigned with an anchor role. This means, tail part anchors the flavin ring into the reaction pocket of LSD1 enzyme and orienting it with lysine residue correctly. In this light, we suggest that in addition to the anchor role, the tail part contribute to the stabilization of stationary points shown in the chemical mechanism by decreasing their energies. By this means, the tail part may decrease the energy barrier of the reaction. However, as the TS structures could not be optimized, making comment will not be healthy.

4.3.2 'Reaction Pocket' System

As the final step for the QM/MM calculations, a system that contains all of the enzyme residues and the reaction pocket species in order to investigate the effect of the neighboring residues on the demethylation reaction. In this system, QM part contains the flavin ring, dimethyl lysine, two crystal water molecules and molecular oxygen, and the MM part contains the tail part of the FAD molecule, 15 H3 tail residues and 22 neighboring enzyme residues that were described in Chapter 3. The charge of the MM part is fixed at the value of +3. This system will be referred to as 'Reaction Pocket' from this point on in this thesis.

Step 1 reactants, i.e. Structure 1, of 'Reaction Pocket' were prepared following the procedure explained in Chapter 3. A view of this system is given in Figure 4.17. In this figure, FAD, dimethyl lysine, two water molecules and molecular oxygen are shown in ball-and-stick representation, and enzyme residues are shown in wireframe representation. In the figure, the hydrogen bonds and Pi-Plus ($\Pi - +$) interactions are also shown. Flavin ring interacts with Lys661 residue via Pi-Plus interactions; while molecular oxygen is stabilized by flavin ring and dimethyl lysine is stabilized by several neighboring enzyme residues through hydrogen bonds.

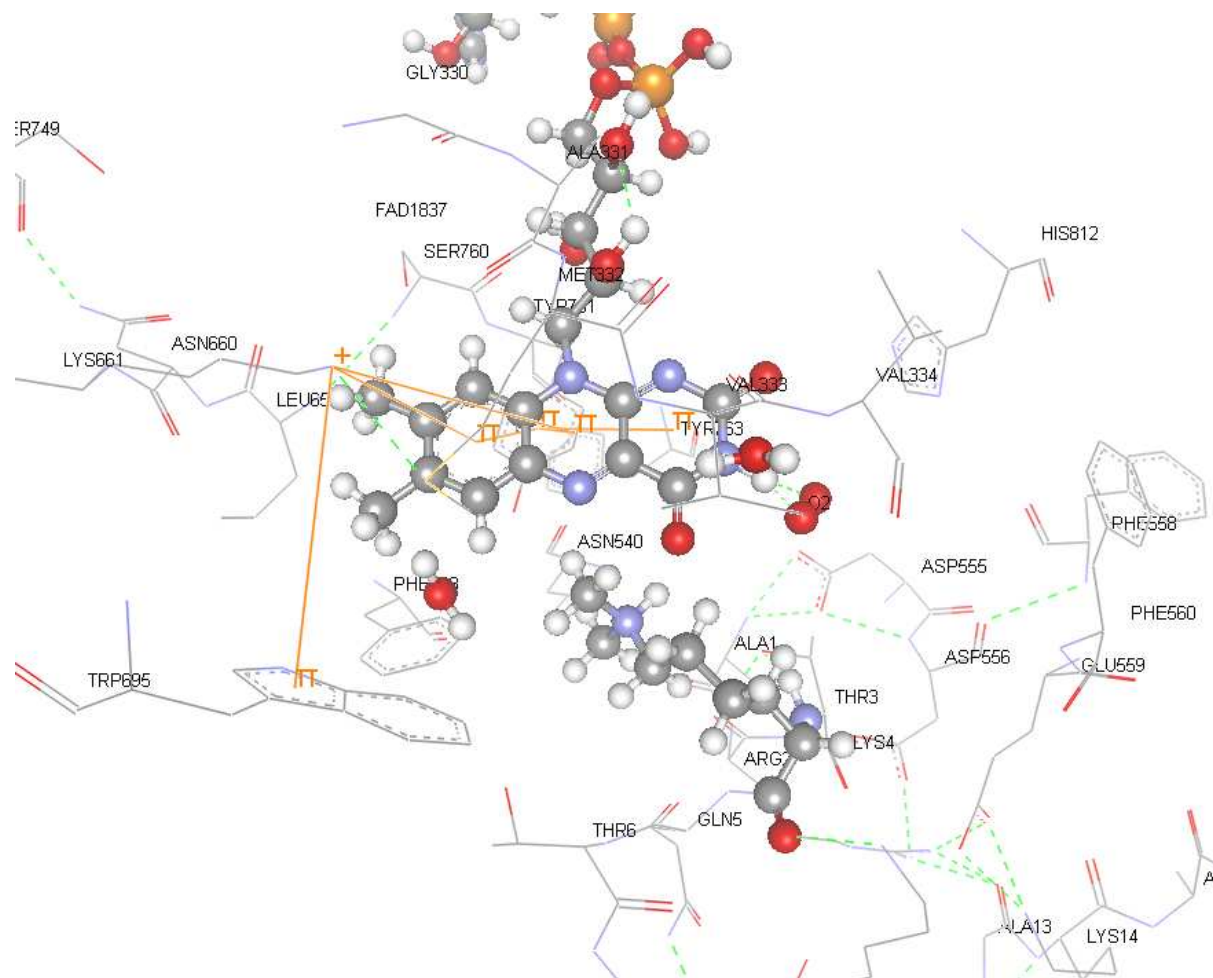


Figure 4. 17: A view from Step 1 reactant of ‘Reaction Pocket’ QM/MM system. (See text for system definition). Reaction pocket species given in ball-and-stick representation. Enzyme residues are given in wireframe representation. Residue names and ID numbers are given as labels. Pi-Plus and hydrogen bonding interactions are shown in solid orange lines and dashed green lines, respectively. Figure is prepared in Accelrys DS Visualizer.

Efforts for optimizing the TS structures for ‘Reaction Pocket’ system also did not succeed, just like the case of the other QM/MM system, i.e. ‘OnlyFAD’ system. Similar to ‘OnlyFAD’ system, geometries optimizations of TS1 and TS2 input structures converged to structures resembling the products of the corresponding steps. The same explanations as given for ‘OnlyFAD’ system can be presented for the failure of geometry optimization of TS structures, since almost the same observations are made. Therefore, there is no need to explain them here again. A difference between two systems is that no optimized geometry that resembles the TS2 structure could be obtained, contrary to the case of ‘OnlyFAD’ system.

Potential and Gibb's free energy profiles are given in Figure 4.18. In the energy profile TS points, i.e. Structure 2 and 4, are missing as explained previously. While preparing the energy profiles, the following multiplicity values were used. Structure 1: quintet; Structure 3: quintet; Structure 5: triplet; Structure 6: singlet; Structure 8: singlet; Structure 10: singlet. All of the structures were optimized only at a single surface at which the optimized geometry has the lowest energy. These surfaces were determined based on the previous observations made in the energy profiles of only-QM systems.

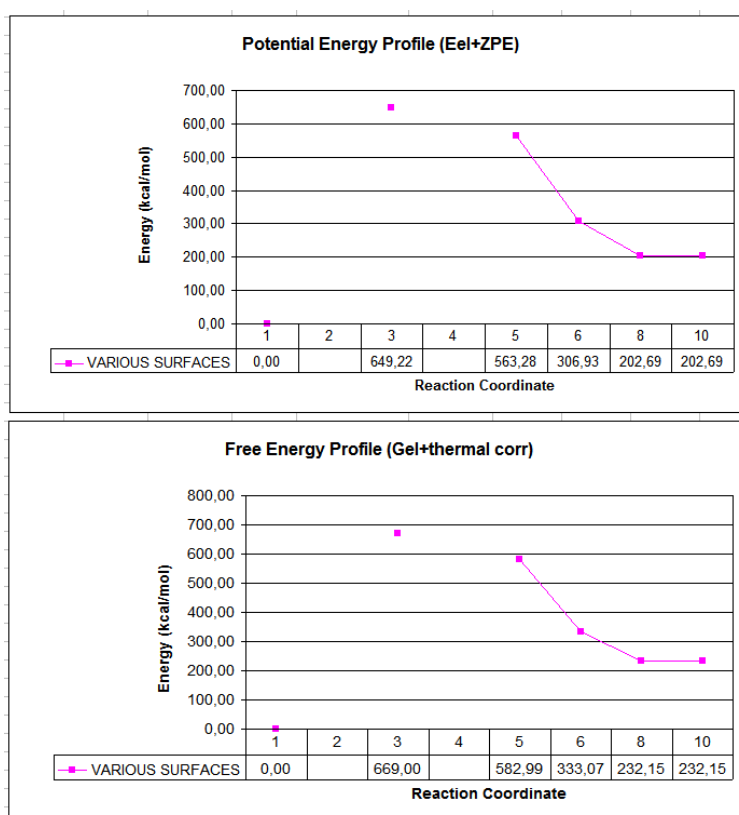


Figure 4. 18: Potential and Gibb's free energy profiles for 'Reaction Pocket' system obtained at B3LYP/6-31g* level. (See text for system definition).

Energy profiles seem to be unreliable in terms of the energy differences between the stationary points. There are 100 to 700 kcal/mol energy differences, which are non-physical and unfeasible. Besides, based on these energy profiles, the reaction seems to be endothermic, which is contradicting with the previous energy profiles belonging to other systems. However, it will be pointless to make use of these energy profiles for making comments, since they show unreliable results.

The non-physical and unreliable results presented in the energy profiles of 'Reaction Pocket' system may be caused of the following reasons.

The geometry optimizations were done at gas phase, even though the enzymatic reaction, i.e. demethylation, is occurring in solvent medium, i.e. water, in real life. Besides, not all of the crystal water molecules, except the two placed very near to flavin ring, are present in the simulation system. This prevents the buffer effect that water provides to stabilize the enzyme residues, H3 tail residues and FAD molecule. Therefore, either all of the crystal water molecules should be added explicitly to the system or an implicit solvent effect should be used. Based on the fact that the use of implicit solvent method and ONIOM method has been very recently developed, use of explicit water molecules will be more reliable. These calculations could not be realized because of the lack of time; however they should be conducted as a future work.

CHAPTER 5

CONCLUSION and FUTURE WORK

Post-translational modifications are involved in many cell-regulatory processes and abnormal levels of these modifications lead to many crucial disorders and diseases. (Detailed information can be found in Chapter 2). Among these modifications, de/methylation process has been relative recently discovered. Methylation occurs on lysine and arginine residues placed on histone tails. Lysine methylation can be of three levels, i.e. mono-, di-, and trimethylated, each of which shows the number of methyl group added to that lysine or arginine residue. Demethylation of these lysine residues are catalyzed either by Jumonji family (for tri- and dimethyl lysines) or LSD family enzymes (for mono- and dimethyl lysines).

In this study, demethylation reaction of fourth lysine residue on H3 tail, i.e. H3K4, which is catalyzed by LSD1, is analyzed in details in order to generate the potential and Gibb's free energy profiles and the thermodynamics and structural properties of the stationary points of the reaction.

For this purpose, as the start-up, a very-detailed, chemically-plausible reaction mechanism is offered by our group for demethylation of a dimethyl lysine (Figure 4.1). The proposed chemical mechanism consists of four steps and ten stationary points. This mechanism is considered to be also valid for monomethyl lysine demethylation. Afterwards, the stationary points shown in this mechanism were modeled using QM and QM/MM hybrid methods.

From the crystal structure and chemical mechanism, it is determined that demethylation reaction occurs between flavin ring of the FAD molecule, which is covalently bound to LSD1, and the methylated lysine. Beside these species, two crystal water molecules and molecular oxygen are also directly involved in the demethylation process. For analyzing the effect of the reaction pocket components on the reaction, seven different simulation systems (five of them are only-QM system and two of them are hybrid QM/MM system) were created. Each of these systems contains some or all of the reaction pocket components that were just mentioned. Using the energy profiles obtained for all of the systems, important deductions were made. The most important ones will be given in the following part.

A trial for obtaining energy profile of the start-up system at semi-empirical PM3 level failed despite our efforts. This shows that as the system contains radical structures as shown in proposed mechanisms, PM3 cannot be successful in simulating this reaction system. Therefore, all of the energy profiles were obtained at higher level methods: B3LYP and MP2.

Looking at the energy profiles prepared for dimethyl and mono methyl lysine demethylation at MP2 level, calculated reaction rates are in agreement with the experimental values in the order of 10^5 and 10^1 , respectively, which are chemically reliable results. Similarly, reaction rate of dimethyl lysine demethylation was shown to be higher compared to mono methyl lysine, which is in agreement with experimental studies. Moreover, the energy profiles show that the reaction is exothermic, therefore energetically favorable. These findings prove that the proposed mechanism is actually suitable for the demethylation of lysines.

While analyzing the chemical mechanism in more detail, we showed in Chapter 4 that during the reaction, system jumps from one energy surface to another by changing spin states as explained in Chapter 3 in order to stabilize the stationary points depicted in the proposed mechanism.

In details, Step 1 was considered and determined as the rate-determining step by having higher energy barrier compared to Step 2. However, switching from triplet surface to quintet surface causes an increase in the energy of Structure 1 (referring to Figure 4.1) by about 90 kcal/mol at MP2-level calculations in dimethyl and monomethyl lysine simulation systems. This spin state change decreases the energy barrier for the first step to about 1 kcal/mol assigning the second step as the rate-determining step. The switch from triplet to quintet surface is not energetically favorable and thus we suggest that this can be only realized in the presence of the enzyme residues.

After Step 2, when the need for the enzyme presence comes to an end, we proposed that the system jumps back to triplet surface and then to singlet surface in a step-wise manner. The MP2-level calculations, which were carried out for checking if this is the case, show that jumping from quintet to triplet surface provides energy stabilization about 70 kcal/mol for Structure 5 (Figure 4.1) of dimethyl and mono methyl systems. Similarly, changing the energy surface from triplet to singlet for Structure 6 provides a stabilization of ca. 90 kcal/mol.

On the other hand, as can be seen from the reaction mechanism (Figure 4.1), lysine is depicted to be migrating to the top of the flavin ring for the second hydrogen transfer between

FAD and lysine to occur. When there is no water or molecular oxygen in the system this migration occurs spontaneously during the optimization. However, when molecular oxygen and two water molecules are present, lysine is stuck beneath the flavin ring during optimization. The input geometries, in which the lysine is pulled to the top of flavin ring manually, yield optimized geometries that have lower energies by about 10 kcal/mol compared to the ones, in which the lysine is placed beneath the flavin ring. Besides, one of the water molecules (preferably the one found at the right-hand side of the flavin ring) accompanies the lysine throughout the migration.

Another observation is that the presence of molecular oxygen and crystal water molecules is not crucial for the reaction, i.e. reaction occurs with an acceptable reaction rate, however when these molecules are present the energy barriers for Step 1 and Step 2 decreases by about 5 kcal/mol at B3LYP level calculations. Therefore, stabilizer role may be assigned to these molecules.

Besides, in Step 3 of the mechanism, a water molecule is consumed in the hydration of aminium cation. In order to determine which one of two water molecules is preferred for this process, we compared the energies of two trials, in each of which one of the water molecules is bound to aminium cation. As the result, it was seen that it is energetically favorable if the water molecule on the right-hand-side of flavin ring is consumed. This is supported by the fact that the position of the right-hand-side water molecule is very close to the aminium cation at the end of Step 2.

Moreover, after Step 2 in the mechanism, FAD is reduced to FADH₂ by taking the hydrogen atoms of the substrate lysine and it should be re-oxidized by molecular oxygen to FAD in order to be able to demethylate another lysine. During this phenomenon, molecular oxygen, i.e. O₂, is reduced to hydrogen peroxide, i.e. H₂O₂. In the chemical mechanism, this process is proposed to be taking place in an intermediary step after the second step. This is supported by the decrease in the energy of the system by 15 and 25 kcal/mol in MP2 level calculations for dimethyl and monomethyl lysines, respectively. Besides, if the reduction of O₂ is not done in one step before the reaction goes on, then there will be radicals, which will destabilize the system.

On the other hand, from the QM/MM calculations, it was observed that the tail part of the FAD molecule, i.e. the molecule without the flavin ring, plays a stabilizer role by lowering

the energy of the stationary points on the mechanism in addition to its previously-shown anchor role for covalent-binding of FAD to LSD1 enzyme.

Nevertheless, calculations for QM/MM systems were not successful as much as the ones for only-QM systems. Transition state structures could not be optimized properly. That means, the optimized geometries do not resemble the corresponding TS structures, instead they resemble the products. This may be caused by the reasons that were explained in Chapter 4.3. More time and effort should be spent for these QM/MM systems in order to obtain more reliable results.

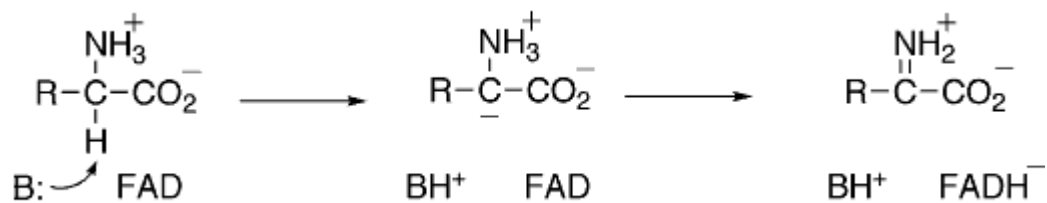
On the other hand, demethylation process occurs in solvent medium, i.e. water, in the cells in real life. Therefore, the reaction path analysis of this process using QM-methods should also be done in the presence of implicit solvent effect in order to make more realistic observations on the demethylation process. There are reliable methods, performance of which has been proved by previous studies, such as PCM for implicit solvent effect.

The analysis study presented in this thesis forms a solid knowledge basis for the demethylation of dimethyl and monomethyl lysines. Based on the observations given in this study, further comparative analysis may be done in the presence of inhibitor molecules. As explained in Chapter 2, there are inhibitor molecule candidates already available in the literature. Analysis of the energy profiles in the presence of these inhibitor molecules may be compared with the ones conducted in this study. At the same time, novel inhibitor candidates may be designed based on the observations presented in this study. In that sense, this study paves the road for investigating the inhibition of lysine demethylation process at molecular level.

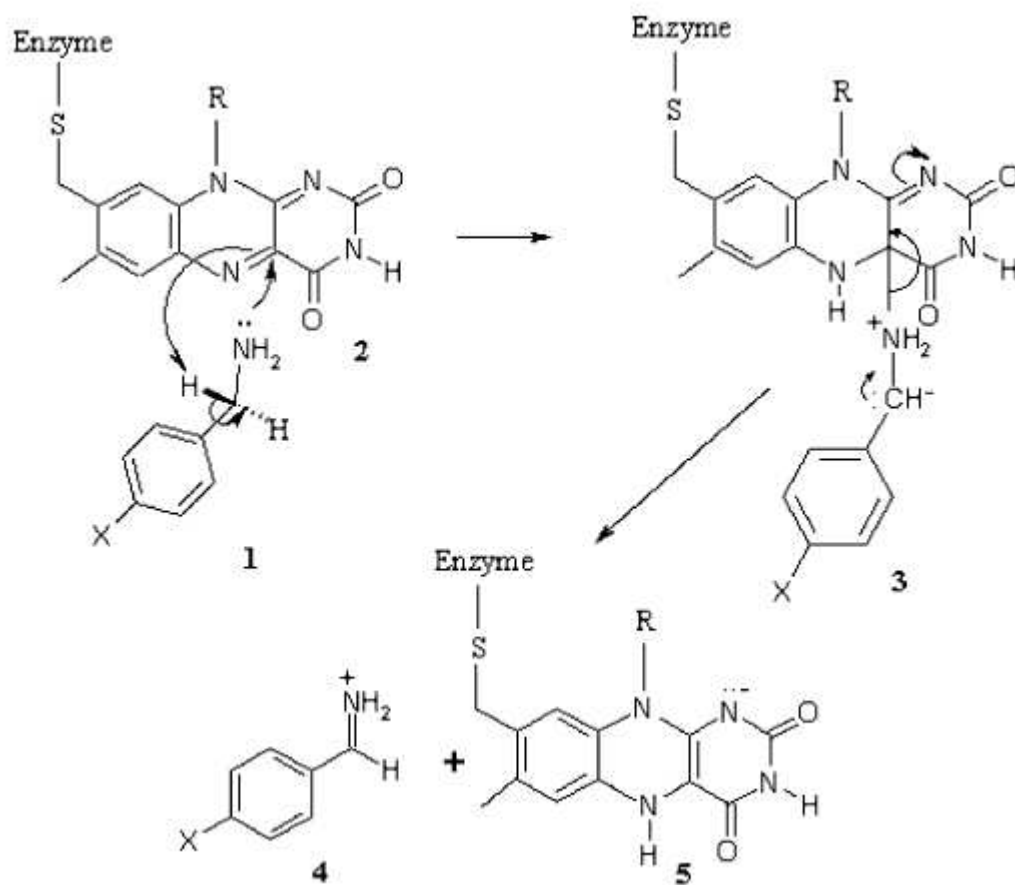
APPENDICES

A – Amine Oxidation Reaction Mechanisms

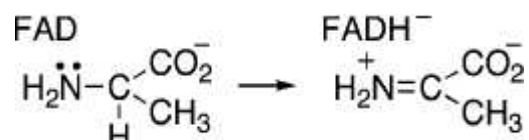
A1 – Carbonion mechanism for amine oxidation [94].



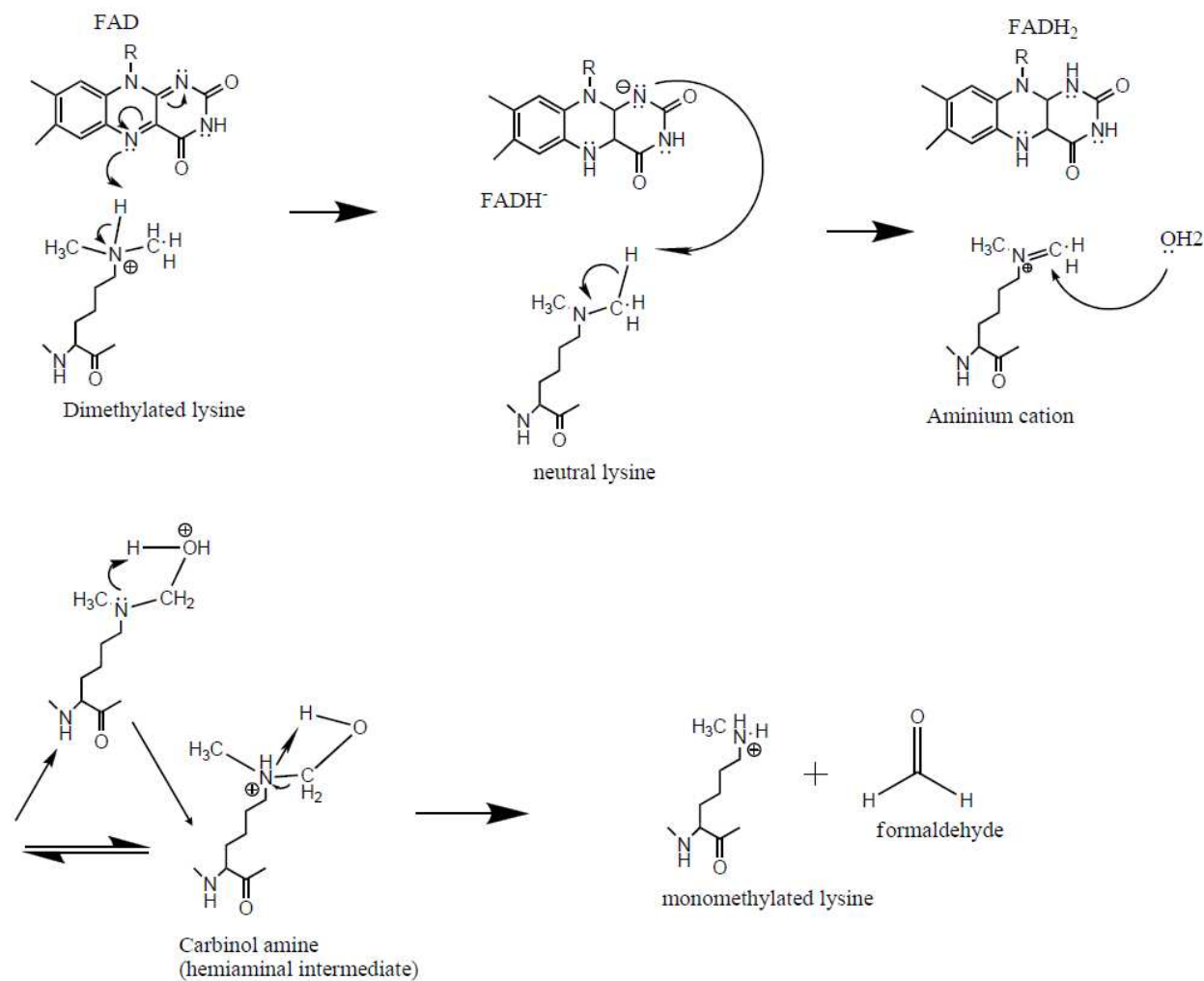
A2- Addition-Elimination mechanism for amine oxidation proposed by Miller and Edmondson [95].



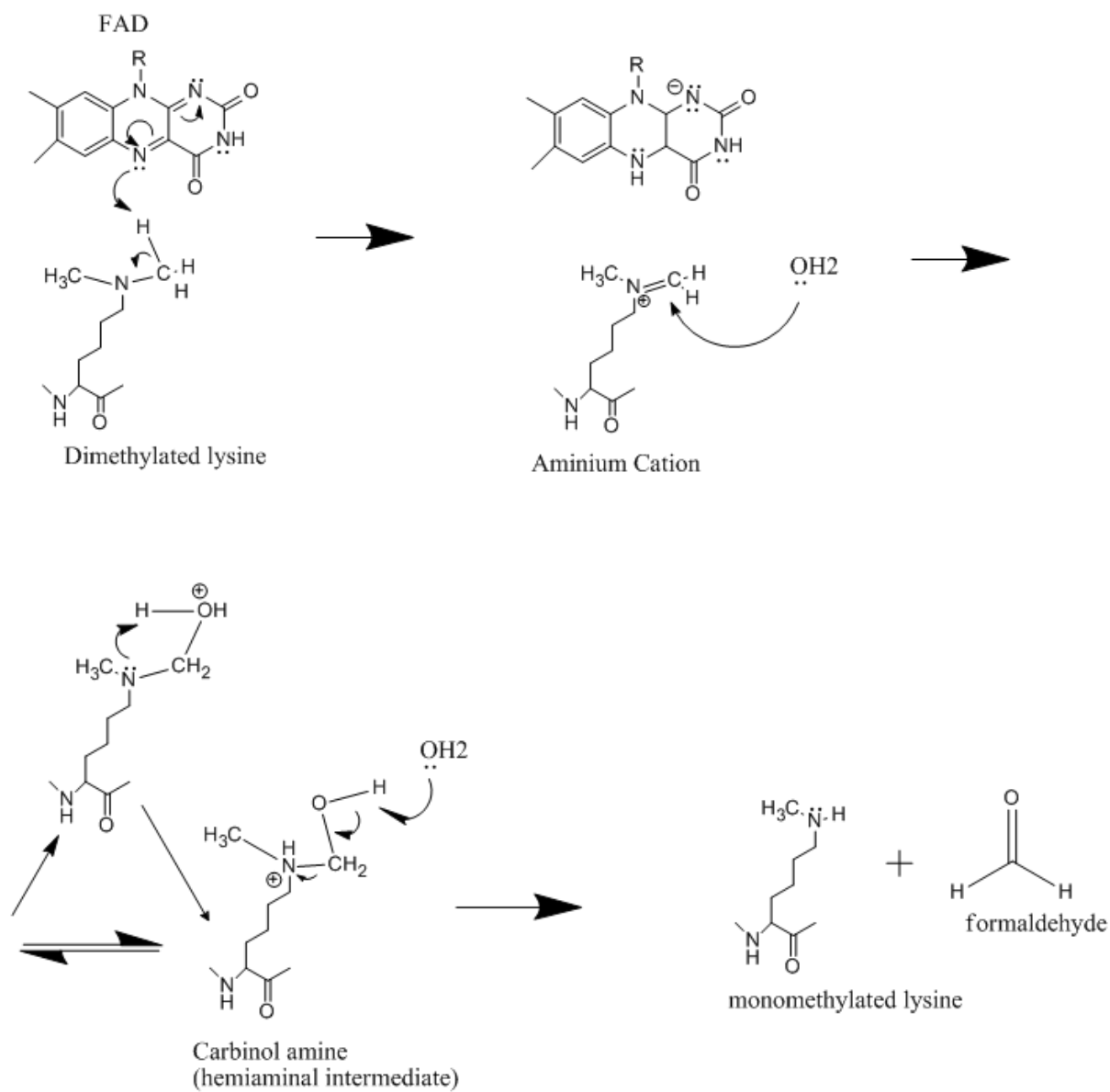
A3 – Hydride Transfer Mechanism proposed by Ralph et al. for TMO catalyzed amine oxidation [97].



A4 – The reaction mechanism (pathway) proposed by us compatible for LSD1 catalyzed demethylation process of positively charged dimethylated lysine.

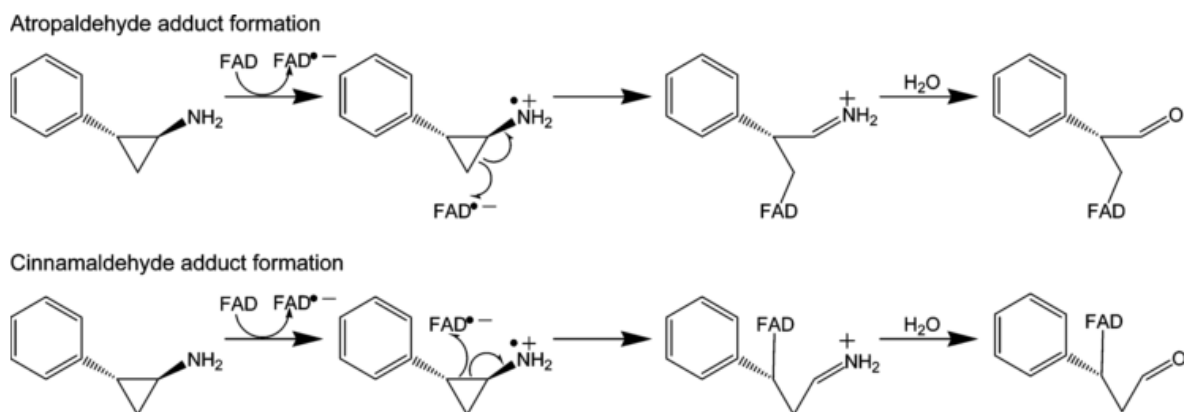


A5 – The reaction mechanism (pathway) proposed by us compatible for LSD1 catalyzed demethylation process of neutral dimethylated lysine.



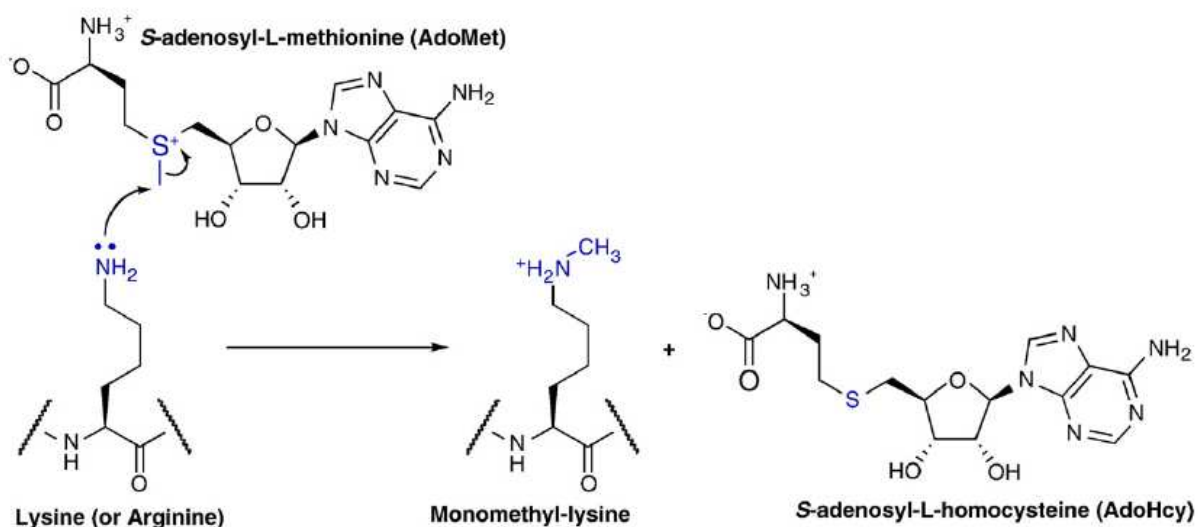
B – Inhibitor Related

B1 – Reaction pathways for the production of tranlycypromine-flavin complex [101].

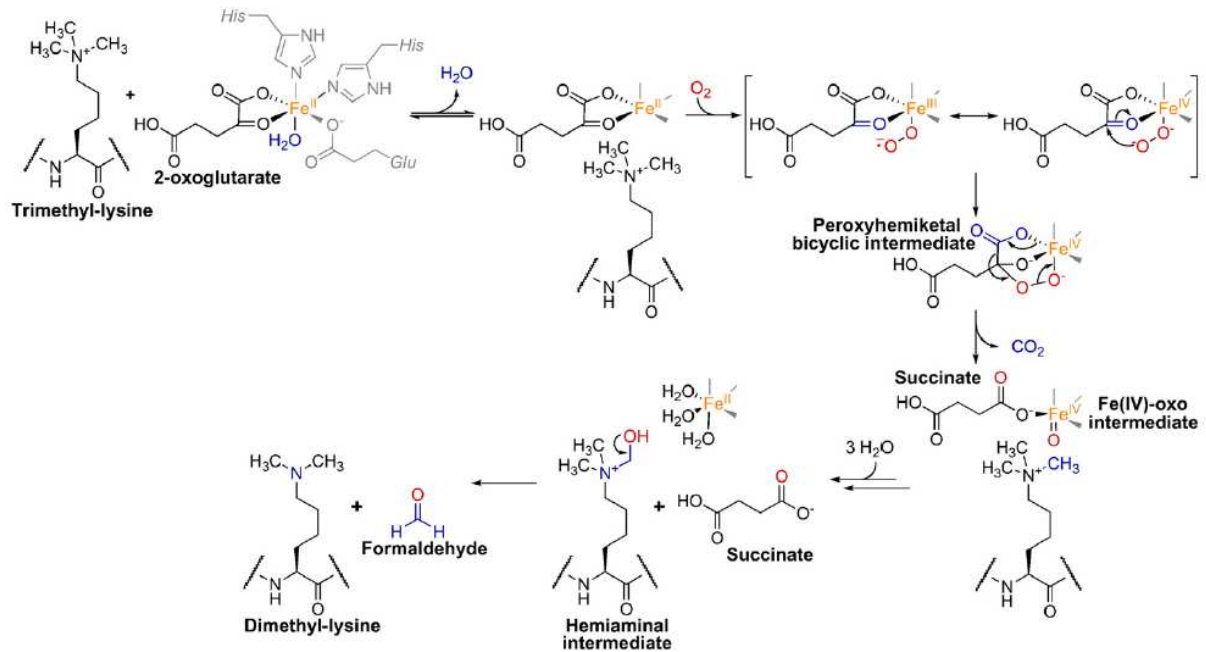


C – Reaction Mechanisms

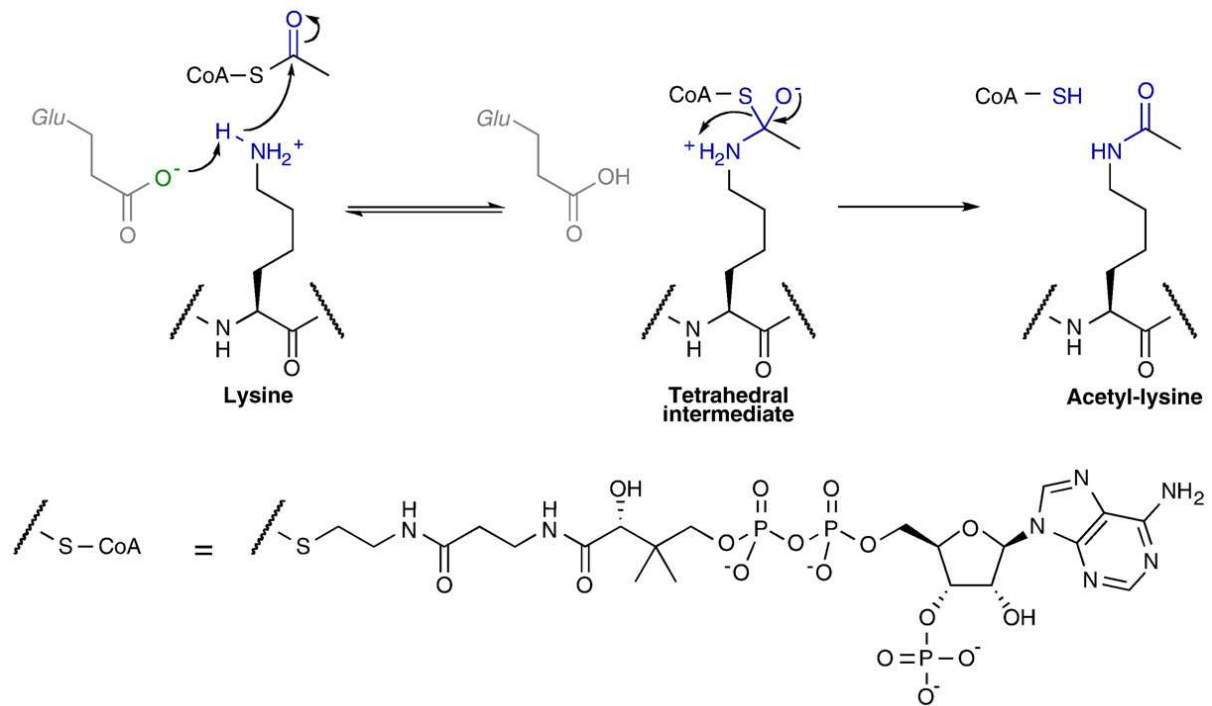
C1 – General chemical mechanism for AdoMet-dependent histone lysine methyl transferases. Set domain, Dot1/KMT4 and arginine methyl transferases use mechanisms similar to this mechanism.



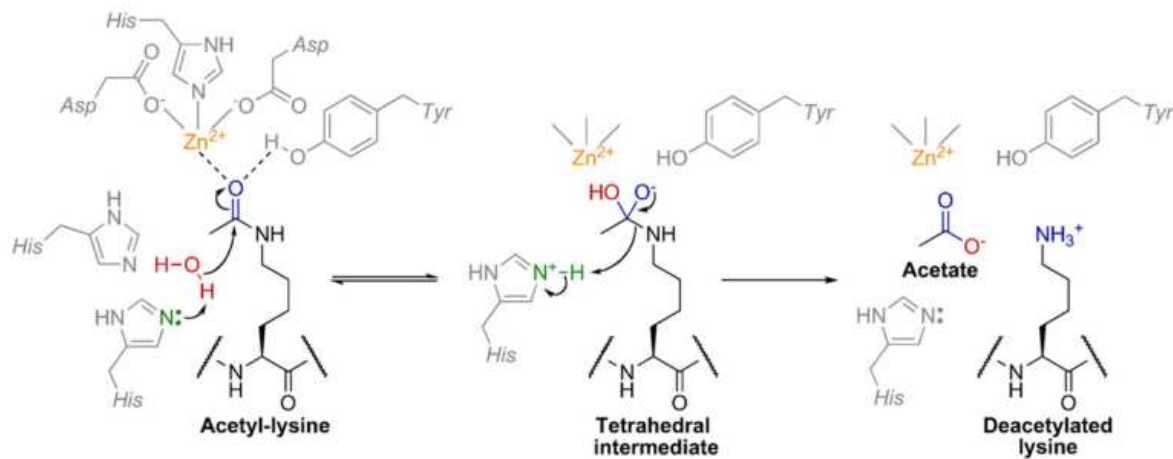
C2 – Proposed reaction (chemical) mechanism for lysine demethylation catalyzed by JHDM (Jumonji Family) enzymes [17].



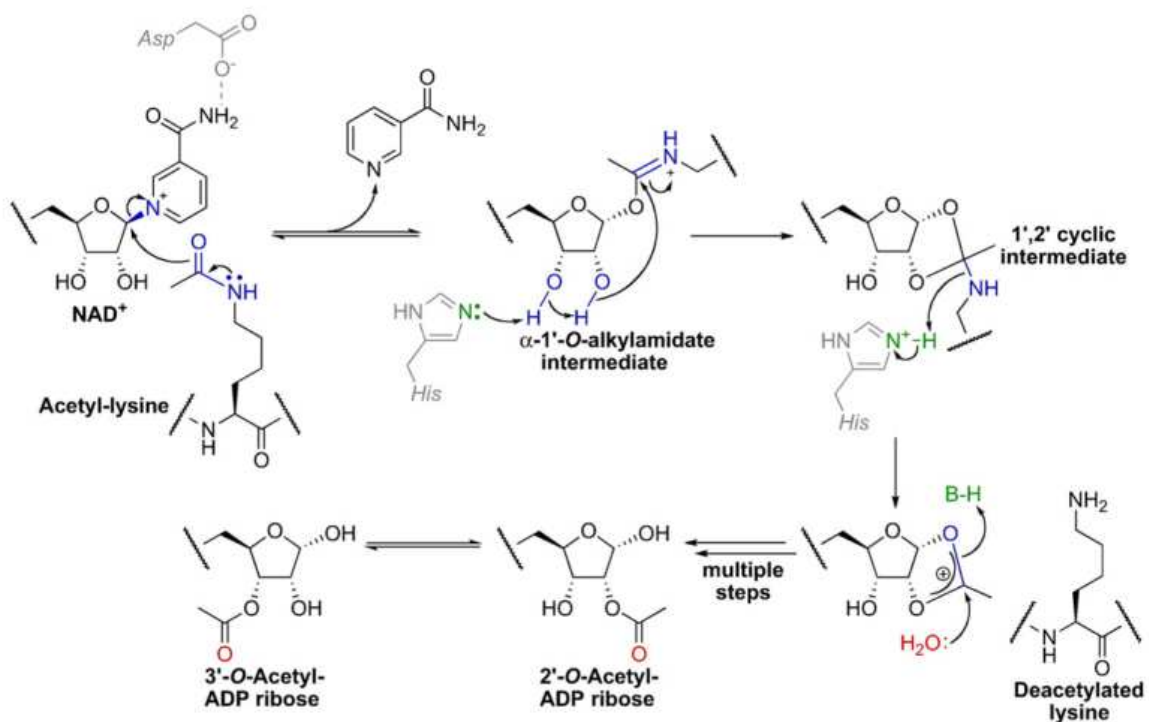
C3 – The general proposed reaction pathway for histone acetyltransferases [17].



C4 – Proposed chemical mechanism for histone lysine deacetylation catalyzed by class I/II/IV HDACs [17].



C5 - Proposed chemical mechanism for histone lysine deacetylation catalyzed by class III HDACs (Sirtuins) [17].



D – Related Tables

D1 - The table of the methylation sites and demethylases that show which demethylase demethylate which demethylation site. (Prepared based on 2007 data) [155].

	LSD1	JHDM1	JHDM2A	JMJD2A	JMJD2B	JMJD2C	JMJD2D
H3K4me3	-	-	-	-	-	-	-
H3K4me2	+	-	-	-	-	-	-
H3K4me1	+	-	-	-	-	-	-
H3K9me3	-	-	-	+	+	+	+
H3K9me2	+	-	+	+/-	-	+	+
H3K9me1	+	-	+	-	-	-	-
H3K27me3	-	-	-	-	-	-	-
H3K27me2	-	-	-	-	-	-	-
H3K27me1	-	-	-	-	-	-	-
H3K36me3	-	-	-	+	-	+	-
H3K36me2	-	+	-	+/-	-	-	-
H3K36me1	-	+	-	-	-	-	-
H3K79me3	-	-	-	-	-	-	-
H3K79me2	-	-	-	-	-	-	-
H3K79me1	-	-	-	-	-	-	-
H3K20me3	-	-	-	-	-	-	-
H3K20me2	-	-	-	-	-	-	-
H3K20me1	-	-	-	-	-	-	-

PS: +/- denotes that corresponding demethylase can demethylate this methylation site but not specific for this site, thus demethylation rate is very low.

D2- The table showing the relationship of arginine methyltransferases with known cancer types. Taken from reference [42], for the references given in parentheses in the table, refer to [42].

Selected arginine methyltransferases and links to cancer

Enzyme	Links to cancer (examples)
PRMT1	Essential component of MLL oncogenic transcriptional complex (Cheung et al., 2007) Coactivator of hormone receptors in hormone dependent cells (Cheng et al., 2004)
PRMT2	Coactivator of the androgen receptor (Meyer et al., 2007)
PRMT4	Essential for estrogen-induced cell cycle progression in breast cancer cells (Frieze et al., 2008)
(CARM1)	Aberrant expression in prostate tumors (Hong et al., 2004) Methylation of CBP contributes to coactivation (Chevallard-Briet et al., 2002) CARM1 knockdown blocks androgen receptor signalling (Majumder et al., 2006)
[5pt] PRMT5 PRMT7	Downregulates the expression of tumor suppressor genes in fibroblasts (Pal et al., 2004). Downregulation sensitizes cancer cells to camptothecin treatment (Verbiest et al., 2008)

D3- The table showing the relationship of lysine methyltransferases with known cancer types. Taken from reference [42], for the references given in parentheses in the table, refer to [42] .

Enzyme	Links to cancer (examples)
SUV39H1/2 (KMT1A/B)	Overexpression in dietary induced tumors (Pogribny et al., 2007) Increased mRNA levels in colon cancer patients (Kang et al., 2007)
G9a (KMT1C)	Contributes to H3K9 dimethylation that is involved in suppressor gene silencing (McGarvey et al., 2006) Knockdown inhibits cancer cell growth (Kondo et al., 2008)
Eu-HMTase1 (KMT1D)	Overexpressed in gland tumors (Aniello et al., 2006)
SETDB1/ESET (KMT1E)	Cooperates with DNA methyltransferase in promotor silencing in tumors (Li et al., 2006)
MLL1 (KMT2A)	Mutations/rearrangements involved in leukemogenesis (Hess, 2004)
MLL4 (KMT2D)	Involved in hepatitis B virus dependent liver carcinogenesis (Saigo et al., 2008)
SMYD2 (KMT3C)	Suppression of p53 transcriptional activity (Huang et al., 2007)
SMYD3	Enhanced breast cancer cell growth due to overexpression (Hamamoto et al., 2006) Overexpression in colorectal and hepatocellular carcinoma (Hamamoto et al., 2004)
DOT1L (KMT4)	Involved in leukemogenesis (Okada et al., 2005)
SET8/PR-SET7 (KMT5A)	Suppresses p53 dependent transcription (Shi et al., 2007)
EZH2 (KMT6)	Associated with aggressive tumor growth in several tumor types (Bachmann et al., 2006) Marker for precancerous state in breast cancerogenesis (Ding et al., 2006) Marker for aggressive breast cancer (Collett et al., 2006) Promotes proliferation and invasiveness of prostate cancer cells (Bryant et al., 2007) Useful as a biomarker for poor prostate cancer prognosis (Cooper et al., 2007)

METHODS RELATED APPENDIX

E- Computational Procedure Related

E1 – Detailed procedure for AutoDock usage.

For the docking of molecular oxygen, the procedure given below is followed.

1. Using *AutoDock Tools 1.5.4*, the PDB files with partial charge and torsion information, i.e. pdbqt, are created for macromolecule, i.e. FAD and neighboring enzyme residues, and ligand, i.e. O₂ or Methylated Lysine.
2. Using *AutoDock Tool*, an input for the docking grid is created. This grid centered on the center of the flavin ring of FAD residue and made spacious enough to cover the entire flavin ring as well as some additional space in which a dimethyl lysine can stay. Energy grid has the spacing of 0.375 Å (approximately one fourth of the length of carbon–carbon covalent bond). The atom types of macromolecule and ligand are defined. After the parameter file is completed, energy grid is created by *AutoGrid* tool of *AutoDock*.
3. Docking parameter files are created using *AutoDock Tools*. Docking is realized using each of the SA, LGA, LS or GA methods, separately, implemented *AutoDock 4.2*. The fine details of the parameter files for each docking type are given below.

Parameters for Genetic Algorithm (GA)

```

...
tran0 random          # initial coordinates/A or random
axisangle0 random    # initial orientation
dihe0 random         # initial dihedrals (relative) or random
tstep 2.0            # translation step/A
qstep 50.0           # quaternion step/deg
dstep 50.0           # torsion step/deg
torsdof 7            # torsional degrees of freedom
rmstol 2.0           # cluster_tolerance/A
extnrg 1000.0        # external grid energy
e0max 0.0 10000      # max initial energy; max number of retries
ga_pop_size 150      # number of individuals in population
ga_num_evals 2500000 # maximum number of energy evaluations
ga_num_generations 27000 # maximum number of generations
ga_elitism 1         # number of top individuals to survive to next generation
ga_mutation_rate 0.02 # rate of gene mutation
ga_crossover_rate 0.8 # rate of crossover
ga_window_size 10    #
ga_cauchy_alpha 0.0  # Alpha parameter of Cauchy distribution
ga_cauchy_beta 1.0   # Beta parameter Cauchy distribution
set_ga               # set the above parameters for GA or LGA
unbound_model bound  # state of unbound ligand
do_global_only 50    # do this many GA runs
analysis             # perform a ranked cluster analysis

```

Parameters for Lamarckian Genetic Algorithm (LGA)

```

...
tran0 random          # initial coordinates/A or random
axisangle0 random    # initial orientation
dihe0 random         # initial dihedrals (relative) or random
tstep 2.0            # translation step/A
qstep 50.0           # quaternion step/deg
dstep 50.0           # torsion step/deg
torsdof 7            # torsional degrees of freedom
rmstol 2.0           # cluster_tolerance/A
extnrg 1000.0        # external grid energy
e0max 0.0 10000      # max initial energy; max number of retries
ga_pop_size 150      # number of individuals in population
ga_num_evals 2500000 # maximum number of energy evaluations
ga_num_generations 27000 # maximum number of generations
ga_elitism 1         # number of top individuals to survive to next generation
ga_mutation_rate 0.02 # rate of gene mutation
ga_crossover_rate 0.8 # rate of crossover
ga_window_size 10    #
ga_cauchy_alpha 0.0  # Alpha parameter of Cauchy distribution
ga_cauchy_beta 1.0   # Beta parameter Cauchy distribution
set_ga               # set the above parameters for GA or LGA
sw_max_its 300       # iterations of Solis & Wets local search
sw_max_succ 4        # consecutive successes before changing rho
sw_max_fail 4        # consecutive failures before changing rho
sw_rho 1.0           # size of local search space to sample
sw_lb_rho 0.01       # lower bound on rho
ls_search_freq 0.06  # probability of performing local search on individual
set_pswl             # set the above pseudo-Solis & Wets parameters
unbound_model bound  # state of unbound ligand
ga_run 100           # do this many hybrid GA-LS runs
analysis             # perform a ranked cluster analysis

```

Parameters for Local Search (LS)

```

...
tran0 random          # initial coordinates/A or random
axisangle0 random    # initial orientation
dihe0 random         # initial dihedrals (relative) or random
tstep 2.0            # translation step/A
qstep 50.0           # quaternion step/deg
dstep 50.0           # torsion step/deg
torsdof 7            # torsional degrees of freedom
rmstol 2.0           # cluster_tolerance/A
extnrg 1000.0        # external grid energy
e0max 0.0 10000      # max initial energy; max number of retries
ga_pop_size 150      # number of individuals in population
sw_max_its 300       # iterations of Solis & Wets local search
sw_max_succ 4        # consecutive successes before changing rho
sw_max_fail 4        # consecutive failures before changing rho
sw_rho 1.0           # size of local search space to sample
sw_lb_rho 0.01       # lower bound on rho
ls_search_freq 0.06  # probability of performing local search on individual
set_psw1             # set the above pseudo-Solis & Wets parameters
unbound_model bound  # state of unbound ligand
do_local_only 500    # do this many LS runs
analysis             # perform a ranked cluster analysis

```

Parameters for Simulated Annealing (SA)

```

...
tran0 random          # initial coordinates/A or random
axisangle0 random    # initial orientation
dihe0 random         # initial dihedrals (relative) or random
tstep 2.0            # translation step/A
qstep 50.0           # quaternion step/deg
dstep 50.0           # torsion step/deg
torsdof 7            # torsional degrees of freedom
rmstol 2.0           # cluster_tolerance/A
extnrg 1000.0        # external grid energy
e0max 0.0 10000      # max initial energy; max number of retries
rt0 1000.0           # initial annealing temperature (times gas constant)
rtrf 0.95            # annealing temperature reduction factor
trnrf 1.0            # per cycle reduction factor for translation
quarf 1.0            # per cycle reduction factor for quaternions
dihrf 1.0            # per cycle reduction factor for dihedrals
runs 100             #
cycles 50            # number of temperature reduction cycles
accs 100             # maximum number of accepted steps per cycle
rejs 100             # maximum number of rejected steps per cycle
select m             # state selection flag: (m)inimum or (l)ast state
unbound_model bound # state of unbound ligand
simanneal            # do as many SA runs as set by runs keyword above
analysis             # perform a ranked cluster analysis

```

E2 – Tricks against convergence problem in *Gaussian* geometry optimizations

- Convergence criteria may be lowered, but this must be done with caution. If two structures are compared, they must be optimized with the same convergence criteria to yield meaningful comparison. Lowering convergence criteria may lead to undesired optimized structures.
- For wave-function convergence problems, *scf=xqc* is a useful tool. It does quadratic optimizations steps, which involves linear steps by default. Convergence is guaranteed but it takes more computational time.
- Especially for TS optimizations, system has an undesired geometry after optimization. We have encountered frequently with this problem, as the chemical reaction involves hydride transfer, which is seen as converging to the reactant or product conformation. For preventing this problem, step-wise optimization can be a useful solution. This procedure involves consecutive optimizations at B3LYP/STO-3G, B3LYP/6-31g and B3LYP/6-31g*, whose input structure is the optimized conformation obtained from lower level calculations.
- By default, an initial guess is created at the beginning of the geometry optimization and updated at every step using information from the previous step. However, creating a new guess for SCF routine at the current step is often useful where wave-function convergence is problematic.
- For geometry optimization convergence problems, decreasing step size may be useful, if the energy surface has steep slopes. Step size is, by default $N \cdot 0.01 \text{ Bohr}$, where N is 30.

BIBLIOGRAPHY

- [1] P. Hu and Y. Zhang, "Catalytic mechanism and product specificity of the histone lysine methyltransferase SET7/9: An ab initio QM/MM-FE study with multiple initial structures," *J. Am. Chem. Soc.*, vol. 128, 2006, pp. 1272-8.
- [2] A. Karytinis, F. Forneris, A. Profumo, and G. Ciossani, "A novel mammalian flavin-dependent histone demethylase," *Journal of Biological*, vol. 284, 2009, pp. 17775-17782.
- [3] J.C. Culhane and P.A. Cole, "LSD1 and the chemistry of histone demethylation," *Current opinion in chemical biology*, vol. 11, 2007, pp. 561-8.
- [4] Y. Shi and J.R. Whetstine, "Dynamic regulation of histone lysine methylation by demethylases.," *Molecular cell*, vol. 25, 2007, pp. 1-14.
- [5] M. Yang, J.C. Culhane, L.M. Szewczuk, P. Jalili, H.L. Ball, M. Machius, P.a. Cole, and H. Yu, "Structural basis for the inhibition of the LSD1 histone demethylase by the antidepressant trans-2-phenylcyclopropylamine.," *Biochemistry*, vol. 46, 2007, pp. 8058-65.
- [6] K. Ahmad and S. Henikoff, "The histone variant H3. 3 marks active chromatin by replication-independent nucleosome assembly," *Molecular cell*, vol. 9, 2002, p. 1191-1200.
- [7] C. David Allis, J. Bowen, G. Abraham, C. Glover, and M. Gorovsky, "Proteolytic processing of histone H3 in chromatin: a physiologically regulated event in *Tetrahymena* micronuclei," *Cell*, vol. 20, 1980, p. 55-64.
- [8] S. Briggs, M. Bryk, B. Strahl, W. Cheung, J. Davie, S. Dent, F. Winston, and C. Allis, "Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in *Saccharomyces cerevisiae*," *Genes & development*, vol. 15, 2001, p. 3286.
- [9] K. Johnson, D. Pflugh, D. Yu, D. Hesslein, K. Lin, A. Bothwell, A. Thomas-Tikhonenko, D. Schatz, and K. Calame, "B cell-specific loss of histone 3 lysine 9 methylation in the VH locus depends on Pax5," *Nature immunology*, vol. 5, 2004, p. 853-861.
- [10] P. Byvoet, G. Shepherd, J. Hardin, and B. Noland, "The distribution and turnover of labeled methyl groups in histone fractions of cultured mammalian cells.," *Archives of biochemistry and biophysics*, vol. 148, 1972, p. 558.

-
- [11] J. Duerre and C. Lee, "In vivo methylation and turnover of rat brain histones," *Journal of Neurochemistry*, vol. 23, 1974, pp. 541-547.
- [12] S. Nielsen, R. Schneider, U. Bauer, A. Bannister, A. Morrison, D. O'Carroll, R. Firestein, M. Cleary, T. Jenuwein, R. Herrera, and others, "Rb targets histone H3 methylation and HP1 to promoters," *Nature*, vol. 412, 2001, p. 561-565.
- [13] B. Strahl, S. Briggs, C. Brame, J. Caldwell, S. Koh, H. Ma, R. Cook, J. Shabanowitz, D. Hunt, M. Stallcup, and others, "Methylation of histone H4 at arginine 3 occurs in vivo and is mediated by the nuclear receptor coactivator PRMT1," *Current Biology*, vol. 11, 2001, p. 996-1000.
- [14] U. Bauer, S. Daujat, S.J. Nielsen, K. Nightingale, and T. Kouzarides, "Methylation at arginine 17 of histone H3 is linked to gene activation.," *EMBO reports*, vol. 3, 2002, pp. 39-44.
- [15] W. Paik and S. Kim, "Enzymatic demethylation of calf thymus histones," *Biochemical and Biophysical Research*, vol. 51, 1973, pp. 781-788.
- [16] Y. Shi, F. Lan, C. Matson, P. Mulligan, J. Whetstine, and PA, "Histone demethylation mediated by the nuclear amine oxidase homolog LSD1," *Cell*, vol. 119, 2004, pp. 941-953.
- [17] B. Smith and J. Denu, "Chemical mechanisms of histone lysine and arginine modifications," *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, vol. 1789, 2009, p. 45-57.
- [18] A.J. Mulholland, "Computational enzymology: modelling the mechanisms of biological catalysts.," *Biochemical Society transactions*, vol. 36, 2008, pp. 22-6.
- [19] H. Guo and H. Guo, "Mechanism of histone methylation catalyzed by protein lysine methyltransferase SET7/9 and origin of product specificity.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, 2007, pp. 8797-802.
- [20] S. Erdem, Ö. Karahan, İ. Yıldız, and K. Yelekçi, "A computational study on the amine-oxidation mechanism of monoamine oxidase: Insight into the polar nucleophilic mechanism," *Organic & Biomolecular*, vol. 4, 2006, p. 646-658.
- [21] K. Luger, A. Mader, R. Richmond, D. Sargent, and T. Richmond, "Crystal structure of the nucleosome core particle at 2.8 Å resolution," *Nature*, vol. 389, 1997, p. 251-260.
- [22] P.B. Talbert and S. Henikoff, "Histone variants--ancient wrap artists of the epigenome.," *Nature reviews. Molecular cell biology*, vol. 28, 2010, pp. 264-75.
- [23] "Nucleosome," 2010, p. <http://en.wikipedia.org/wiki/Nucleosome>.

- [24] E. Metzger and R. Schüle, "The expanding world of histone lysine demethylases," *Nature Structural & Molecular Biology*, 2007.
- [25] C. Wood, A. Snijders, J. Williamson, C. Reynolds, J. Baldwin, and M. Dickman, "Post-translational modifications of the linker histone variants and their association with cell mechanisms.," *The FEBS journal*, vol. 276, 2009, pp. 3685-97.
- [26] Y. Fan, T. Nikitina, J. Zhao, T. Fleury, and R. Bhattacharyya, "Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation," *Cell*, vol. 123, 2005, pp. 1199-1212.
- [27] E. Bradbury, R. Inglis, and H. Matthews, "Control of cell division by very lysine rich histone (F1) phosphorylation," *Nature*, vol. 247, 1974, p. 257-261.
- [28] B. Sarg, W. Helliger, H. Talasz, F. Barbara, and H.L. Herbert, "HH (2006) Histone H1 phosphorylation occurs site-specifically during interphase and mitosis: identification of a novel phosphorylation site on histone H1," *J Biol Chem*, vol. 281, 2005, pp. 6573-6580.
- [29] S. Santaguida and A. Musacchio, "The life and miracles of kinetochores," *The EMBO Journal*, vol. 28, 2009, pp. 2511-2531.
- [30] J. Wiśniewski, A. Zougman, S. Krüger, and M. Mann, "Mass spectrometric mapping of linker histone H1 variants reveals multiple acetylations, methylations, and phosphorylation as well as differences between cell culture and tissue," *Molecular & Cellular Proteomics*, vol. 6, 2007, p. 72.
- [31] J. Couture and R.C. Trievel, "Histone-modifying enzymes: encrypting an enigmatic epigenetic code.," *Current opinion in structural biology*, vol. 16, 2006, pp. 753-60.
- [32] A. Kuzmichev, T. Jenuwein, P. Tempst, and D. Reinberg, "Different EZH2-containing complexes target methylation of histone H1 or nucleosomal histone H3," *Molecular cell*, vol. 14, 2004, p. 183-193.
- [33] M.G. Lee, C. Wynder, D.a. Bochar, M. Hakimi, N. Cooch, and R. Shiekhattar, "Functional interplay between histone demethylase and deacetylase enzymes.," *Molecular and cellular biology*, vol. 26, 2006, pp. 6395-402.
- [34] N. Cervoni and M. Szyf, "Demethylase activity is directed by histone acetylation.," *The Journal of biological chemistry*, vol. 276, 2001, pp. 40778-87.
- [35] J. Rice and C. Allis, "Histone methylation versus histone acetylation: new insights into epigenetic regulation," *Current Opinion in Cell Biology*, vol. 13, 2001, p. 263-273.
- [36] R. Marmorstein and R.C. Trievel, "Histone modifying enzymes: structures, mechanisms, and specificities.," *Biochimica et biophysica acta*, vol. 1789, 2009, pp. 58-68.

- [37] J. Taunton, C.A. Hassig, and S.L. Schreiber, "A mammalian histone deacetylase related to the yeast regulator Rpd3p," vol. 272, 1996, pp. 408-411.
- [38] A. Bannister, R. Schneider, and T. Kouzarides, "Histone Methylation: Dynamic or Static?," *Cell*, vol. 109, 2002, pp. 801-6.
- [39] P. Trojer and D. Reinberg, "Histone lysine demethylases and their impact on epigenetics.," *Cell*, vol. 125, 2006, pp. 213-7.
- [40] F. Forneris, C. Binda, E. Battaglioli, and A. Mattevi, "LSD1: oxidative chemistry for multifaceted functions in chromatin regulation.," *Trends in biochemical sciences*, vol. 33, 2008, pp. 181-9.
- [41] J. Schneider and A. Shilatifard, "Histone demethylation by hydroxylation: chemistry in action.," *ACS chemical biology*, vol. 1, 2006, pp. 75-81.
- [42] A. Spannhoff, W. Sippl, and M. Jung, "Cancer treatment of the future: inhibitors of histone methyltransferases.," *The international journal of biochemistry & cell biology*, vol. 41, 2009, pp. 4-11.
- [43] P. Byvoet, "In vivo turnover and distribution of radio-N-methyl in arginine-rich histones from rat tissues.," *Archives of biochemistry and biophysics*, vol. 152, 1972, pp. 887-888.
- [44] T. Borun, D. Pearson, and W. Paik, "Studies of histone methylation during the HeLa S-3 cell cycle," *Journal of Biological Chemistry*, vol. 247, 1972, p. 4288-4298.
- [45] A. Annunziato, M. Eason, and C. Perry, "Relationship between methylation and acetylation of arginine-rich histones in cycling and arrested HeLa cells," *Biochemistry*, vol. 34, 1995, p. 2916-2924.
- [46] G. Cuthbert, S. Daujat, A. Snowden, and H. Erdjument-, "Histone deimination antagonizes arginine methylation," *Cell*, vol. 118, 2004, pp. 545-553.
- [47] S. Kim, L. Benoiton, and W. Paik, "Epsilon-Alkyllysine: Purification and Properties of the Enzyme.," *The Journal of biological chemistry*, vol. 239, 1964, pp. 3790-3796.
- [48] W. Paik and S. Kim, "Alkyllysine new assay method, purification, and biological significance," *Archives of Biochemistry and Biophysics*, vol. 165, 1974, pp. 369-378.
- [49] M. Lachner, D. O'Carroll, S. Rea, K. Mechtler, and T. Jenuwein, "Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins.," *Nature*, vol. 410, 2001, pp. 116-20.
- [50] A.J. Bannister and T. Kouzarides, "Reversing histone methylation.," *Nature*, vol. 436, 2005, pp. 1103-6.

- [51] R.K. McGinty, J. Kim, C. Chatterjee, R.G. Roeder, and T.W. Muir, "Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation.," *Nature*, vol. 453, 2008, pp. 812-6.
- [52] C. Niehrs, "Active DNA demethylation and DNA repair," *Differentiation*, vol. 77, 2009, pp. 1-11.
- [53] J. He, Q. Yang, and L. Chang, "Dynamic DNA methylation and histone modifications contribute to lentiviral transgene silencing in murine embryonic carcinoma cells," *Journal of virology*, vol. 79, 2005, p. 13497–13508.
- [54] J. Zhu and X. Yao, "Use of DNA methylation for cancer detection: promises and challenges.," *The international journal of biochemistry & cell biology*, vol. 41, 2009, pp. 147-54.
- [55] L. Méndez-Acuña, M.D. Tomaso, Palitti F., and W. Martinez-Lopez, "Histone Post-Translational Modifications in DNA Damage Response," *Cytogenetic and*, vol. 128, 2010, pp. 28-36.
- [56] A. Bode and Z. Dong, "Post-translational modification of p53 in tumorigenesis," *Nature Reviews Cancer*, vol. 4, 2004, pp. 793-805.
- [57] B. Pérez-Cadahía, B. Drobic, and JR, "H3 phosphorylation: dual role in mitosis and interphase," *Biochemistry and Cell*, vol. 709, 2009, pp. 695-709.
- [58] C.L. Brooks and W. Gu, "Ubiquitination, phosphorylation and acetylation: the molecular basis for p53 regulation," *Current Opinion in Cell Biology*, vol. 15, 2003, pp. 164-171.
- [59] D. Ribet and P. Cossart, "Post-translational modifications in host cells during bacterial infection.," *FEBS letters*, vol. 584, 2010, pp. 2748-58.
- [60] S.L. Berger, "Histone modifications in transcriptional regulation.," *Current opinion in genetics & development*, vol. 12, 2002, pp. 142-8.
- [61] K. Zhang and S.Y. Dent, "Histone modifying enzymes and cancer: going beyond histones.," *Journal of cellular biochemistry*, vol. 96, 2005, pp. 1137-48.
- [62] T. Jiang, X. Zhou, K. Taghizadeh, and M, "N-formylation of lysine in histone proteins as a secondary modification arising from oxidative DNA damage," *Proceedings of the*, vol. 104, 2007, pp. 60-65.
- [63] C.L. Peterson and M. Laniel, "Histones and histone modifications.," *Current biology : CB*, vol. 14, 2004, pp. R546-51.

- [64] P. Chi, C.D. Allis, and G.G. Wang, "Covalent histone modifications--miswritten, misinterpreted and mis-erased in human cancers.," *Nature reviews. Cancer*, vol. 10, 2010, pp. 457-69.
- [65] Y. Shi, C. Matson, F. Lan, S. Iwase, T. Baba, and Y. Shi, "Regulation of LSD1 histone demethylase activity by its associated factors.," *Molecular cell*, vol. 19, 2005, pp. 857-64.
- [66] P. Stavropoulos, G. Blobel, and A. Hoelz, "Crystal structure and mechanism of human lysine-specific demethylase-1," *Nature structural & molecular*, vol. 13, 2006, pp. 626-632.
- [67] R. Silverman, S. Hoffman, and W. Catus, "A mechanism for mitochondrial monoamine oxidase catalyzed amine oxidation," *Journal of the American*, vol. 102, 1980, pp. 7126-7128.
- [68] Y. Chen, Y. Yang, F. Wang, K. Wan, K. Yamane, Y. Zhang, and M. Lei, "Crystal structure of human histone lysine-specific demethylase 1 (LSD1).," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, 2006, pp. 13956-61.
- [69] G. Da, J. Lenkart, K. Zhao, and R, "Structure and function of the SWIRM domain, a conserved protein module found in chromatin regulatory complexes," *Proceedings of the*, vol. 103, 2006, p. 2057-2062.
- [70] J. Huang, R. Sengupta, A. Espejo, M. Lee, and J. Dorsey, "p53 is regulated by the lysine demethylase LSD1," *Nature*, vol. 449, 2007, pp. 105-109.
- [71] M.G. Lee, C. Wynder, N. Cooch, and R. Shiekhattar, "An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation.," *Nature*, vol. 437, 2005, pp. 432-5.
- [72] M. Yang, C.B. Gocke, X. Luo, D. Borek, D.R. Tomchick, M. Machius, Z. Otwinowski, and H. Yu, "Structural basis for CoREST-dependent demethylation of nucleosomes by the human LSD1 histone demethylase.," *Molecular cell*, vol. 23, 2006, pp. 377-87.
- [73] M. Hakimi, D. Bochar, and J. Chenoweth, "A core-BRAF35 complex containing histone deacetylase mediates repression of neuronal-specific genes," *Proceedings of the*, vol. 99, 2002, p. 7420-7425.
- [74] Y. Tsukada, J. Fang, H. Erdjument-Bromage, M.E. Warren, C.H. Borchers, P. Tempst, and Y. Zhang, "Histone demethylation by a family of JmjC domain-containing proteins.," *Nature*, vol. 439, 2006, pp. 811-6.
- [75] Y. Huang, J. Fang, M.T. Bedford, Y. Zhang, and R. Xu, "Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A.," *Science (New York, N.Y.)*, vol. 312, 2006, pp. 748-51.

- [76] S. Ng, K. Kavanagh, M. McDonough, D. Butler, and ES, "Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity," *Nature*, vol. 448, 2007, pp. 87-91.
- [77] Z. Chen, J. Zang, J. Kappler, X. Hong, F. Crawford, Q. Wang, F. Lan, C. Jiang, J. Whetstone, S. Dai, K. Hansen, Y. Shi, and G. Zhang, "Structural basis of the recognition of a methylated histone tail by JMJD2A.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, 2007, pp. 10818-23.
- [78] J. Couture, E. Collazo, P.a. Ortiz-Tello, J.S. Brunzelle, and R.C. Trievel, "Specificity and mechanism of JMJD2A, a trimethyllysine-specific histone demethylase.," *Nature structural & molecular biology*, vol. 14, 2007, pp. 689-95.
- [79] M. Wissmann, N. Yin, J. Müller, H. Greschik, and BD, "Cooperative demethylation by JMJD2C and LSD1 promotes androgen receptor-dependent gene expression," *Nature cell*, vol. 9, 2007, pp. 347-353.
- [80] T. Takeuchi, Y. Watanabe, T. Takano-Shimizu, and S. Kondo, "Roles of jumonji and jumonji family genes in chromatin regulation and development.," *Developmental dynamics : an official publication of the American Association of Anatomists*, vol. 235, 2006, pp. 2449-59.
- [81] R. Schneider, A. Bannister, and T. Kouzarides, "Unsafe SETs: histone lysine methyltransferases and cancer," *Trends in biochemical*, vol. 27, 2002, p. 396-402.
- [82] M. Glozak and E. Seto, "Histone deacetylases and cancer," *Oncogene*, vol. 26, 2007, p. 5420-5432.
- [83] X. Chen, F. Niroomand, Z. Liu, A. Zankl, and H. Katus, "Expression of nitric oxide related enzymes in coronary heart disease," *Basic research in*, vol. 101, 2006, p. 346-353.
- [84] A. Suzuki, R. Yamada, and K. Yamamoto, "Citrullination by peptidylarginine deiminase in rheumatoid arthritis," *Annals of the New York Academy of Sciences*, vol. 1108, 2007, p. 323-339.
- [85] J. Milne and J. Denu, "The Sirtuin family: therapeutic targets to treat diseases of aging," *Current opinion in chemical biology*, 2008.
- [86] V. Longo and B. Kennedy, "Sirtuins in aging and age-related disease," *Cell*, vol. 126, 2006, p. 257-268.
- [87] T. Anekonda and P. Reddy, "Neuronal protection by sirtuins in Alzheimer's disease," *Journal of neurochemistry*, vol. 95, 2006, pp. 305-313.
- [88] M. Bedford and S. Richard, "Arginine Methylation:: An Emerging Regulator of Protein Function," *Molecular cell*, vol. 18, 2005, p. 263-272.

- [89] H. Hong, C. Kao, M. Jeng, J. Eble, M. Koch, and TA, "Aberrant expression of CARM1, a transcriptional coactivator of androgen receptor, in the development of prostate carcinoma and androgen-independent status," *Cancer*, vol. 101, 2004, pp. 83-89.
- [90] S. Majumder, Y. Liu, O.F. III, J. Mohler, and YE, "Involvement of arginine methyltransferase CARM1 in androgen receptor function and prostate cancer cell viability," *Prostate*, vol. 66, 2006, p. 1292–1301.
- [91] B. Chang, Y. Chen, Y. Zhao, and R. Bruick, "JMJD6 is a histone arginine demethylase," *Science*, vol. 318, 2007, pp. 444-447.
- [92] R.B. Silverman, "Radical Ideas about Monoamine Oxidase," *Accounts of Chemical Research*, vol. 28, 1995, pp. 335-342.
- [93] N.S. Scrutton, "Chemical aspects of amine oxidation by flavoprotein enzymes.," *Natural product reports*, vol. 21, 2004, pp. 722-30.
- [94] P.F. Fitzpatrick, "Carbanion versus hydride transfer mechanisms in flavoprotein-catalyzed dehydrogenations.," *Bioorganic chemistry*, vol. 32, 2004, pp. 125-39.
- [95] J.R. Miller and D.E. Edmondson, "Structure-activity relationships in the oxidation of para-substituted benzylamine analogues by recombinant human liver monoamine oxidase A.," *Biochemistry*, vol. 38, 1999, pp. 13670-83.
- [96] R. Silverman, "Radical thoughts about the life of MAO," *Progress in Brain Research*, vol. 106, 1995, p. 23–31.
- [97] E.C. Ralph, M.a. Anderson, W.W. Cleland, and P.F. Fitzpatrick, "Mechanistic studies of the flavoenzyme tryptophan 2-monooxygenase: deuterium and ¹⁵N kinetic isotope effects on alanine oxidation by an L-amino acid oxidase.," *Biochemistry*, vol. 45, 2006, pp. 15844-52.
- [98] M. Giorgio, M. Trinei, E. Migliaccio, and PG, "Hydrogen peroxide: a metabolic by-product or a common mediator of ageing signals?," *Nature Reviews Molecular*, vol. 8, 2007, p. 722– 728.
- [99] E. TYIHÁK, L. Trezl, and B. Szende, "Formaldehyde cycle and the phases of stress syndrome," *of the New York Academy of*, vol. 851, 1998, p. 259–270.
- [100] Y. Huang, E. Greene, and T.M. Stewart, "Inhibition of lysine-specific demethylase 1 by polyamine analogues results in reexpression of aberrantly silenced genes," *Proceedings of the*, vol. 104, 2007, p. 8023–8028.
- [101] D.M. Schmidt and D.G. McCafferty, "trans-2-Phenylcyclopropylamine is a mechanism-based inactivator of the histone demethylase LSD1.," *Biochemistry*, vol. 46, 2007, pp. 4408-16.

- [102] M. Lee, C. Wynder, D. Schmidt, D. McCafferty, and R. Shiekhattar, "Histone H3 lysine 4 demethylation is a target of nonselective antidepressive medications," *Chemistry & biology*, vol. 13, 2006, p. 563–567.
- [103] M.G. Lee, C. Wynder, D.M. Schmidt, D.G. McCafferty, and R. Shiekhattar, "Histone H3 lysine 4 demethylation is a target of nonselective antidepressive medications.," *Chemistry & biology*, vol. 13, 2006, pp. 563-7.
- [104] S. Mimasu, T. Sengoku, S. Fukuzawa, and T. Umehara, "Crystal structure of histone demethylase LSD1 and tranlylcypromine at 2.25 Å," *Biochemical and*, vol. 366, 2008, pp. 15-22.
- [105] M. Yang, J. Culhane, L. Szewczuk, and CB, "Structural basis of histone demethylation by LSD1 revealed by suicide inactivation," *Nature Structural &*, vol. 14, 2007, pp. 535-539.
- [106] J.C. Culhane, L.M. Szewczuk, X. Liu, G. Da, R. Marmorstein, and P.a. Cole, "A mechanism-based inactivator for histone demethylase LSD1.," *Journal of the American Chemical Society*, vol. 128, 2006, pp. 4536-7.
- [107] C. Binda, F. Hubalek, M. Li, Y. Herzig, J. Sterling, and DE, "Crystal structures of monoamine oxidase B in complex with four inhibitors of the N-propargylaminoindan class," *J. Med.*, 2004.
- [108] K. Yelekçi, O. Karahan, and M. Toprakçi, "Docking of novel reversible monoamine oxidase-B inhibitors: efficient prediction of ligand binding sites and estimation of inhibitors thermodynamic properties.," *Journal of neural transmission (Vienna, Austria : 1996)*, vol. 114, 2007, pp. 725-32.
- [109] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Jr., J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, O. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, and D.J. Fox, "Gaussian 09, Revision A.02," 2009.
- [110] F. Jensen, *Introduction to computational chemistry*, Chichester, England: John Wiley & Sons, 1999.
- [111] W. Hehre, *A guide to molecular mechanics and quantum chemical calculations*, Irvine, CA, USA: Wavefunction Inc., 2003.

- [112] A. Warshel, "Computer simulations of enzyme catalysis: methods, progress, and insights.," *Annual review of biophysics and biomolecular structure*, vol. 32, 2003, pp. 425-43.
- [113] J. Foresman and Æ. Frisch, *Exploring chemistry with electronic structure methods*, Pittsburgh, PA, USA: Gaussian Inc., 1996.
- [114] R. Poli and J.N. Harvey, "Spin forbidden chemical reactions of transition metal compounds. New ideas and new computational challenges," *Chemical Society Reviews*, vol. 32, 2003, pp. 1-8.
- [115] J.N. Harvey, "Understanding the kinetics of spin-forbidden chemical reactions.," *Physical chemistry chemical physics : PCCP*, vol. 9, 2007, pp. 331-43.
- [116] D. Bakowies and W. Thiel, "Hybrid models for combined quantum mechanical and molecular mechanical approaches," *The Journal of Physical Chemistry*, vol. 100, 1996, p. 10580–10594.
- [117] T. Vreven, K. Byun, I. Komáromi, S. Dapprich, J.a. Montgomery, K. Morokuma, and M. Frisch, "Combining quantum mechanics methods with molecular mechanics methods in ONIOM," *J. Chem. Theory*, vol. 2, 2006, pp. 815-826.
- [118] A. Warshel and R. Weiss, "An empirical valence bond approach for comparing reactions in solutions and in enzymes," *Journal of the American Chemical Society*, vol. 102, 1980, p. 6218–6226.
- [119] N. Reuter, A. Dejaegere, B. Maignet, and M. Karplus, "Frontier Bonds in QM/MM Methods: A Comparison of Different Approaches," *The Journal of Physical Chemistry A*, vol. 104, 2000, pp. 1720-1735.
- [120] V. Théry, D. Rinaldi, J. Rivail, and B, "Quantum mechanical computations on very large molecular systems: The local self-consistent field method," *Journal of*, 1994.
- [121] M. Svensson, S. Humbel, R.D. Froese, T. Matsubara, S. Sieber, and K. Morokuma, "ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels–Alder Reactions and Pt(P(t -Bu) 3) 2 + H 2 Oxidative Addition," *The Journal of Physical Chemistry*, vol. 100, 1996, pp. 19357-19363.
- [122] X. Li and M.J. Frisch, "Energy-Represented Direct Inversion in the Iterative Subspace within a Hybrid Geometry Optimization Method," *Journal of Chemical Theory and Computation*, vol. 2, 2006, pp. 835-839.
- [123] K.N. Kudin, G.E. Scuseria, and E. Cancès, "A black-box self-consistent field convergence algorithm: One step closer," *The Journal of Chemical Physics*, vol. 116, 2002, p. 8255.

- [124] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Jr., J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, O. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, and D.J. Fox, "Gaussian 09 User's Reference," http://www.gaussian.com/g_tech/g_ur/g09help.htm, 2010.
- [125] J. Harris, "Simplified method for calculating the energy of weakly interacting fragments," *Physical Review B*, vol. 31, 1985, pp. 1770-1779.
- [126] K. Fukui, "The path of chemical reactions-the IRC approach," *Accounts of Chemical Research*, vol. 14, 1981, pp. 363-368.
- [127] J. Cioslowski, "A new population analysis based on atomic polar tensors," *Journal of the American Chemical Society*, vol. 111, 1989.
- [128] J.a. Montgomery, M.J. Frisch, J.W. Ochterski, and G.a. Petersson, "A complete basis set model chemistry. VII. Use of the minimum population localization method," *The Journal of Chemical Physics*, vol. 112, 2000, p. 6532.
- [129] G. Morris, G. Goodshell, R. Huey, W. Hart, R. Halliday, R. Belew, and A. Olson, "Autodock (version 3.05)," 1999.
- [130] B. Unal, A. Gursoy, and B. Erman, "Inhibitor Peptide Design for NF- κ B: Markov Model & Genetic Algorithm," *HIBIT 2010 Proceedings*, Antalya: IEEE Xplore, 2010, pp. 72-80.
- [131] G. Morris, D. Goodsell, and R. Halliday, "Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function," *Journal of Computational Chemistry*, vol. 19, 1998, pp. 1639-1662.
- [132] F. Forneris, C. Binda, A. Adamo, E. Battaglioli, and A. Mattevi, "Structural basis of LSD1-CoREST selectivity in histone H3 recognition.," *The Journal of biological chemistry*, vol. 282, 2007, pp. 20070-4.
- [133] W.L. Jorgensen, D.S. Maxwell, and J. Tirado-Rives, "Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids," *Journal of the American Chemical Society*, vol. 118, 1996, pp. 11225-11236.

- [134] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, and H.J. Berendsen, "GROMACS: fast, flexible, and free.," *Journal of computational chemistry*, vol. 26, 2005, pp. 1701-18.
- [135] H. Xie, Y. Ding, and C. Sun, "Radical Reaction C₃H+NO : A Mechanistic Study," *Journal of Computational Chemistry*, vol. 27, 2006, pp. 641-660.
- [136] S. Xu, "A combined DFT and SCRF study of solvent effects on 4-methyl-3-penten-2-one," *Journal of Molecular Structure: THEOCHEM*, vol. 459, 1999, pp. 163-170.
- [137] G. Cavelier and L.M. Amzel, "Mechanism of NAD(P)H:quinone reductase: Ab initio studies of reduced flavin.," *Proteins*, vol. 43, 2001, pp. 420-32.
- [138] A. Rappe, C. Casewit, K. Colwell, and WA, "UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations," *J. Am. Chem. Soc.*, vol. 2, 1992, pp. 10024-10035.
- [139] X. Lu, F. Tian, Y. Feng, X. Xu, N. Wang, and Q. Zhang, "Sidewall Oxidation and Complexation of Carbon Nanotubes by Base-Catalyzed Cycloaddition of Transition Metal Oxide: A Theoretical Prediction," *Nano Letters*, 2002, pp. 4-6.
- [140] K. Bobuatong, "Effects of the zeolite framework on the adsorption of ethylene and benzene on alkali-exchanged zeolites: an ONIOM study," *Applied Catalysis A: General*, vol. 253, 2003, pp. 49-64.
- [141] J. Lomratsiri, M. Probst, and J. Limtrakul, "Structure and adsorption of a basic probe molecule on H-ZSM-5 nanostructured zeolite: An embedded ONIOM study," *Journal of Molecular Graphics and*, vol. 25, 2006, pp. 219-225.
- [142] E. Derat, "On the link atom distance in the ONIOM scheme. An harmonic approximation analysis," *Journal of Molecular Structure: THEOCHEM*, vol. 632, 2003, pp. 61-69.
- [143] S. Kasuriya, "Adsorption of ethylene, benzene, and ethylbenzene over faujasite zeolites investigated by the ONIOM method," *Journal of Catalysis*, vol. 219, 2003, pp. 320-328.
- [144] Y. Shigeta, "Hybrid QM/MM studies on energetics of malonaldehyde in condensed phase," *International Journal of Quantum Chemistry*, vol. 96, 2004, pp. 32-41.
- [145] D. Pantazis, J. McGrady, F. Maseras, and M, "Critical role of the correlation functional in DFT descriptions of an agostic niobium complex," *J. Chem. Theory*, 2007, pp. 1329-1336.
- [146] J.E. Norton and J. Brédas, "Polarization energies in oligoacene semiconductor crystals.," *Journal of the American Chemical Society*, vol. 130, 2008, pp. 12377-84.

- [147] B. Jansang, T. Nanok, and J. Limtrakul, "Structures and reaction mechanisms of cumene formation via benzene alkylation with propylene in a newly synthesized ITQ-24 zeolite: an embedded ONIOM study.," *The journal of physical chemistry. B*, vol. 110, 2006, pp. 12626-31.
- [148] R. Hart and P.J. O'Malley, "A quantum mechanics/molecular mechanics study of the tyrosine residue, Tyr(D), of Photosystem II.," *Biochimica et biophysica acta*, vol. 1797, 2010, pp. 250-4.
- [149] A. Lodola, M. Mor, J.C. Hermann, G. Tarzia, D. Piomelli, and A.J. Mulholland, "QM/MM modelling of oleamide hydrolysis in fatty acid amide hydrolase (FAAH) reveals a new mechanism of nucleophile activation.," *Chemical communications (Cambridge, England)*, 2005, pp. 4399-401.
- [150] F. Claeysens, K.E. Ranaghan, F.R. Manby, J.N. Harvey, and A.J. Mulholland, "Multiple high-level QM/MM reaction paths demonstrate transition-state stabilization in chorismate mutase: correlation of barrier height with transition-state stabilization.," *Chemical communications (Cambridge, England)*, 2005, pp. 5068-70.
- [151] F. Claeysens, J.N. Harvey, F.R. Manby, R.a. Mata, A.J. Mulholland, K.E. Ranaghan, M. Schütz, S. Thiel, W. Thiel, and H. Werner, "High-accuracy computation of reaction barriers in enzymes.," *Angewandte Chemie (International ed. in English)*, vol. 45, 2006, pp. 6856-9.
- [152] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, and P.a. Kollman, "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules," *Journal of the American Chemical Society*, vol. 117, 1995, pp. 5179-5197.
- [153] S.L. Mayo, B.D. Olafson, and W.a. Goddard, "DREIDING: a generic force field for molecular simulations," *The Journal of Physical Chemistry*, vol. 94, 1990, pp. 8897-8909.
- [154] B. Karasulu, O. Keskin, and B. Erman, "Reaction path analysis for demethylation process of histone tail lysine residues," *5th International Symposium on Health Informatics and Bioinformatics (HIBIT)*, Antalya: IEEE XPlore, 2010, pp. 197-205.
- [155] X. Tian and J. Fang, "Current Perspectives on Histone Demethylases," *Acta Biochimica et Biophysica Sinica*, vol. 39, 2007, pp. 81-88.

VITA

Bora Karasulu was born in Istanbul, Turkey on November, 28 1986. He is an alumnus of Huseyin Avni Sozen Anadolu Lisesi, Istanbul. He received his Bachelor of Science degree in Chemistry from Koç University, Istanbul, in June 2008.

From 2008 to 2010, he was a research and teaching assistant in the Computational Science and Engineering Department of Koç University. He graduated from M.Sc. Program offered at this department. His research interest includes reaction path analysis of histone tail lysine residue demethylation using only-QM and hybrid QM/MM methods.