A Comparative Molecular Dynamics Study of Methylation State Specificty of JMJD2A

by

Özlem Ulucan

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

Koç University
August 2009

Koç University

Graduate School of Sciences and Engineering

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

Özlem Ulucan

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.

_____

Attila Gürsoy, Ph. D.

Date: _____

ABSTRACT

Specific patterns of post-translational modifications of histones act as a molecular "code" recognized and used by non-histone proteins to regulate specific chromatin functions. K9 methylation on Histone 3 (H3) tail, mainly trimethylation, induces formation of constitutive heterochromatin via a well-known pathway, which employs heterochromatin formation protein (HP1) and DNA methyl transferase (DNMT). Jumonji domain containing 2A (JMJD2A) is a histone demethylase that specifically removes K9 and K36 trimethyl marks on H3 tail. This enzyme does not function on monomethyl marks and has almost 20-fold reduced activity on dimethyl forms compared to trimethyl forms.

In order to gain insight into how JMJD2A discriminates between its substrates, we performed molecular dynamics simulations of mono-, di- and trimethylated histone tails in complex with JMJD2A catalytic domain and analyzed positional fluctuations, located the hydrogen bonds and calculated some critical distances. We revealed the importance of water molecules and the oxygen-enclosed environment in appropriate orientation of methylammonium head in the active site. We also calculated binding free energy and energy contribution of each residue. We found out that recognition is mostly driven by van der Waals and Coulombic interactions in enzyme-substrate interface. We also revealed the role of Arg8 on the H3 tail in binding and stabilizing the necessary conformation of substrate peptide.

ÖZET

Histonların translasyon sonrası modifikasyonları ile oluşan moleküler kod, hücre içerisinde histon olmayan proteinler tarafından okunmakta ve kromatinin işlevinin düzenlenmesinde kullanılmaktadır. Histon 3 kuyruğunun 9. lizininin metilasyonu, çoğunlukla trimetilasyonu, heterokromatin oluşum proteini 1'in (HP1) ve DNA metil transferaz'ın (DNMT) da görev aldığı iyi bilinen bir yolak aracılığıyla heterokromatin oluşumunu ve devamlılığını tetikler. Jumonji domeni içeren 2A (JMJD2A), H3 kuyruğu üzerindeki trimetillenmiş lizin9 ve lizin36'nın metil gruplarını özgün olarak uzaklaştıran bir demetilazdır. Bu enzim monometillenmiş lizinler üzerinde hiçbir aktivite göstermezken dimetillenmiş lizinlerde trimetillenmişlere göre 20 kat indirgenmiş bir aktivite gösterir.

Bu enzimin kendi substratlarının metil seviyelerinin ayırtına nasıl vardığını anlamak için, enzimin katalitik domeniyle monometillenmiş, dimetillenmiş ve trimetillenmiş H3 kuyruklarının komplekslerinin moleküler dinamik (MD) simülasyonlarını yaptık. Simülasyon sonrasında bazı atomların konumsal dalgalanmalarının analizini yaptık, hidrojen bağlarını tespit ettik ve bazı kritik uzaklıkları hesapladık. Metilamonyum başının aktif yerde uygun bir biçimde konumlanmasında su moleküllerinin ve oksijenlerce oluşturulmuş cebin önemini ortaya koyduk. Konuyu ayrıca enerji açısından ele aldık; bağlanma enerjisini ve bağlanma enerjisine her aminoasidin katkısını hesapladık. Bulgular bağlanmanın büyük oranda van der Waals ve Coulumb etkileşimlerince sağlandığını ortaya koymakta. Ayrıca 8. arjininin substrat peptidin bağlanmasında ve stabilitesinin sağlanmasındaki önemini gösterdik.

ACKNOWLEDGEMENTS

First of all I would like to thank my advisors Prof. Burak Erman, Assoc. Prof. Özlem Keskin and Assoc. Prof Attila Gürsoy for guidance, help and support throughout this thesis. Working with you has been instructive and exiting.

For financial support I thank to The Scientific and Technological Research Council of Turkey (TÜBITAK).

Work would not have been as pleasant without my project mates. Special thanks to Musa Özboyacı and Deniz Şanlı.

I would also like to thank my office mates M. Müge Karaman, Ömer Kırkağaçlıoğlu, Nuray Dindar and S. Özgür Oğuz and my colleagues Özge Engin and E. Besray Ünal. Thank you very much for helping me to develop my skills and providing your kind friendship any time.

I would like to show my gratitude to Assist. Prof. Mehmet Sayar, Assoc. Prof Alper T. Erdoğan and Assoc Prof. Alper Demir for helping me to develop my skills. I would especially thank to Assoc. Prof. Serdar Kuyucak for helping me to extend my knowledge on Molecular Dynamics.

I owe my deepest gratitude to my family always waiting for me to come home. Thank you for unconditional support!

My special thanks goes to Fesih for his unflagging love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

JMJD2A	Jumonji domain containing 2A
H3K9(me3)	Histone 3 trimethylated at lysine 9
H3K9(me2)	Histone 3 dimethylated at lysine 9
H3K9(me1)	Histone 3 monomethylated at lysine 9
NOG	N-oxalylglycine
Fe(II)	Divalent ferrous cation
α-KG	Alpha ketogluterate
NPT	Constant atom number, volume and temperature
PME	Particle Mesh Ewald
RESP	Restraint Electrostatic Potential
GAFF	General Amber Force Field
ΔG	Binding free energy
MM-PBSA	Molecular mechanics Poisson-Boltzmann surface area
MM-GBSA	Molecular mechanics generalized Born surface area
∇	Gradient
P	Momentum
E_{total}	Total energy
$\overline{G}_{molecule}$	Free energy of the molecule
$\overline{E}_{MM}$	Total mechanical energy in gas phase
$\overline{E}_{bond}$	Bond energies
$\overline{E}_{angle}$	Angular energies
$\overline{E}_{tors}$	Torsional energies

$\overline{E}_{vdW}$	van der Waals energies
$\overline{E}_{elec}$	Columbic energies
$\overline{G}_{PBSA}$	Solvation free energy
$\overline{G}_{non-elec}$	Non-polar salvation energy
$SASA$	Solvent accessible surface area
$H_{molecule}$	Enthalpic contribution of the molecule
TS_{MM}	Entropic contribution
S_{trans}	Translational entropy
S_{rot}	Rotational entropy
S_{vib}	Vibrational entropy
$\Delta\Delta G_{binding}$	Binding free energy difference
RMSD	Root mean square deviation
WT	wild type

Chapter 1

INTRODUCTION

The main aim of my thesis has been to contribute to how the histone demethylase enzyme Jumonji domain containing 2A (JMJD2A) discriminates between methylation states. The main tool of this work has been full atomistic molecular dynamics (MD). MD is widely used to reveal the relationship between molecular structure, movement and function. My ambition with this work has been to apply MD to reveal the relation between the structure and methylation state selectivity of the enzyme JMJD2A. In order to achieve this, MD simulations of JMJD2A catalytic domain in complex with H3 tail mono-, di- and trimethylated at K9 were performed.

JMJD2A specifically removes methyl groups of trimethyllysines at certain postions on N-terminal tail of Histone 3 (H3). Although JMJD2A exhibits reduced activity on dimethyllysine cases, has no activity on monomethyllysine cases. The enzyme began to grow in importance when its overexpression was first detected in certain cancer types. Thereafter the enzyme has become object of interest in many ways by many research groups.

Chen at al. determined the structure of the catalytic-core in complex with methylated Histone 3 Lys 36 (H3K36) peptide substrates[1]. In their study they mainly addressed the sequence specificity of the enzyme. Additionally they found out that the interactions between enzyme and substrate peptide were mainly main chain-main chain interactions. They also assessed the detailed interactions between methyllysine head and its binding

environment. Their claim was that the specificity for a certain methyl group was affected by space and the electrostatic environment of the catalytic center.

Ng at al. probed how JMJD2A discriminates between the methylation states and achieves sequence specificity via resolving JMJD2A catalytic domain in complex with tri-, mono- and trimethyl forms of H3K9 and trimethyl form of H3K36[2]. They proposed a mechanism for how JMJD2A achieves methylation state selectivity involving contribution of water molecules.

Right after that, Couture at al. reported the crystal structure of JMJD2A catalytic-core domain in complex with mono-, di- and trimethylated forms of H3K9 peptide[3]. In their work they stated that the network of C-H---O[4] type hydrogen bonds coordinates the trimethyllysine in JMJD2A active site and positions one methyl group into close proximity of Fe(II).

This thesis consists of five chapters. The first chapter mentions how-and-why of this study. In the present chapter the aim of this work is stated and a brief description of previous works that closely related to the object of interest is given. In the second chapter a proper acknowledgement of previous works is done to derive a sophisticated understanding of the context under a set of subtitles. The third chapter consists of explanations of computational methods that used in this study. In this chapter we deal with parameterization, MD concepts and algorithms, binding free energy calculations with MMPBSA method, entropy calculations, binding free decomposition with MMGBSA method and computational alanine scanning method. In the forth chapter we present our results under a number of titles and subtitles. Here we evaluate our findings and try to relate them with previous works. At the end, we draw conclusions from present work.

Chapter 2

LITERATURE REVIEW

2.1 Some Concepts in Epigenetics

The term “epigenetic” refers to changes in gene expression that are stable over rounds of cell division, and sometimes between generations, but do not involve changes in the underlying DNA sequence of the organism[5].

2.1.1. Chromatin Structure

In eukaryotic cells genomic DNA is packed into the nucleus as chromatin. Unlike the traditional view that considers chromatin as a DNA-packaging device, chromatin represents an additional level of regulation for all metabolic processes that are related to DNA such as replication, repair and gene expression. In this sense chromatin works as a platform where biological signals combine and molecular responses take place.

Chromatin is comprised of DNA, histone proteins and non-histone proteins. The fundamental building block of chromatin is the nucleosome, which consists of 147 bp DNA wrapped around an octamer of histone proteins. Each nucleosome core contains two copies of evolutionarily conserved H2A, H2B, H3 and H4 proteins. Those histone proteins have a globular C-terminal domain that is required for nucleosome formation and a flexible N-terminal tail that projects out from the nucleosome core. Nucleoseomes are then further packed with linker histones (H1) and other structural proteins into higher-order chromatins[6].

Chromatin basically has two different levels of structure: Heterochromatin and euchromatin. Heterochromatin is the portion of the genome that remains packed during interphase. Heterochromatic regions are rich in repetitive sequences, poor in gene content, transcriptionally silent and generally replicate late. On the other hand euchromatin is the rest, which loosens after the transition from metaphase to interphase, includes almost all genes, is transcriptionally active and replicates early. Additionally the structure of chromatin may change transiently due to the presence of cellular signals.

2.1.2. Histone Post-translational Modifications and "the Histone Code"

Histone tails that protrude from nucleosome are subject to a large number of modifications that include methylation, acetylation, ubiquitilation and phosphorilation. These modifications can generate or stabilize binding sites for regulatory proteins, such as transcription factors, proteins in chromatin packing or DNA repair. On the contrary, these modifications may also have counter effects; disrupting or blocking binding sites on chromatin.

Histone modifications occur at multiple and specific sites, which generate various combinations. Accordingly, some modifications co-exist and work cooperatively but are incompatible with others inside the same nucleosome. Methylation of Lysine 4 H3 (H3K4), acethylation of Lysine 14 H3 (K14 H3) and phosphorilation of Serine 10 H3 (H3S10) properly exemplify that case. All three modifications are involved in transcription activation and incompatible with the well-known inhibitory H3 Lysine 9 methylation (H3K9)[7].

Moreover, the biological outcomes of a particular modification may also be influenced by the degree of the modification. Lysine side chains may be mono-, di-, and trimethylated, whereas arginine side chains may be monomethylated or symmetrically or asymmetrically dimethylyated. Histone arginine methylation in general associates with gene activation,

while histone lysine methylation causes to either activation or repression depending on the methylation site and state.

2.1.3. Histone 3 Lysine 9 Methylation and Its Effects

Lysine residues (K4, K9, K27 and K36) on histone 3 tail can be mono-, di- and trimethylated. These differentially methylated residues serve as doking sites for diverse effector proteins which function in various physiological responses. Methylation of K4 on H3 tail associates with eucromatic areas and trimethylation stage of this mark specifically increases upon transcriptional activation. In a similar way methylation of K36 on histone H3 tail correlates with transcriptional activation. By contrast with K4 and K36 marks, K27 is a signal for transcriptional repression and maintenance of silenced chromatin.

K9 methylation on H3 tail, mainly trimethylation, induces formation of constitutive heterochromatin via a well-known pathway. In this pathway, H3 K9 residue is methylated via histone methyl transferase (HMT). Thereupon, H3 K9 mark serves as a site to recruit the heterochromatin formation protein (HP1). HP1 induces recruitment of DNA methyl transferase (DNMT) that results in DNA methylation. DNA methylation is a strict signal for heterochromatin establishment (see Figure 1).

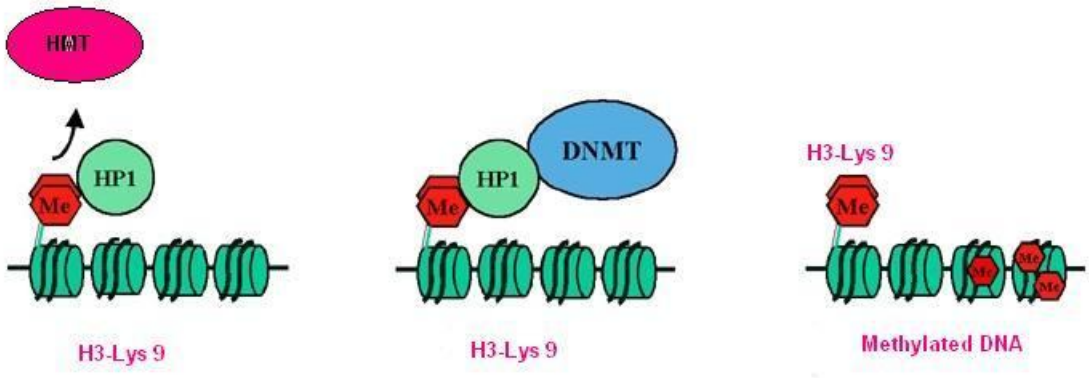

Figure 2.1: Histone H3 Lys 9 induced heterochromatin establishment [8].

A reduction in HP1 in cells results in kinetochore defects, loss of chromosome cohesion and condensation and aberrant chromosome segregation[9]. In addition to these abnormalities, centromere and telomere functions are also impaired. Taking into account the various roles of HP1, it is easy to appreciate the importance of K9 mark which has crucial role in recruitment of HP1.

K9 mono- and dimethylation forms are found to have role in retinoblastoma mediated transcriptional repression of eucromatic genes and in formation of facultative heterochromatin in the inactivated X chromosome of mammalian cells[7].

2.2 Histone Demethylase JMJD2A

2.2.1 Basic Properties

2.2.1.1 Domain Structure

JMJD2A is a histone demethylase that specifically demethlates K9 and K36 trimethyl marks on H3 tail. The whole protein consists of 1064 amino acids which basically form six separate domains: one Jumonji N (JMJN) domain, one Jumonji C (JMJC) domain, two plant homeodomains (PHD) and two tudor domains (see Figure 2).

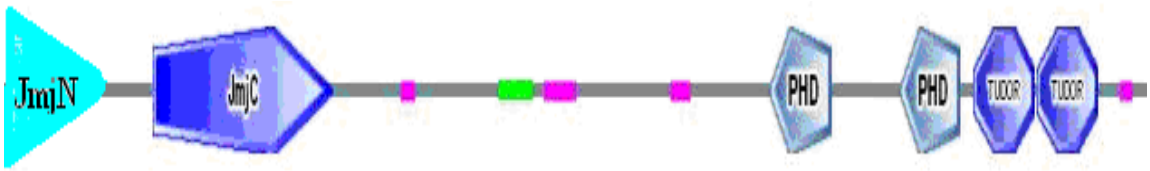

Figure 2.2: Domains of JMJD2A

The catalytic-core domain of JMJD2A enzyme consists of the first 350 amino acids which cover JMJN and JMJC domains at the same time[10].

JMJD2A needs Fe (II) and α-ketoglutarate as cofactors to show demethylation activity [11].

2.2.1.2 Structural Properties

The crystal structure of the catalytic-core domain was firstly determined by Chen at al. in the presence of Fe(II) with and without α-ketoglutarate[10]. In their work they demonstrated that the structure of the catalytic-core domain consists of the JMJN domain, JMJC domain, the C terminal domain and a zinc-finger motif (see Figure 3). Thereafter various crystal structures of JMJD2A catalytic-core domain in the presence of distinct substrates came one after another. Again Chen at al. determined the structure of the catalytic-core complexed with methylated H3K36 peptide substrates[1], Ng at al. brought to light how JMJD2A discriminates between the methylation states and achieves sequence specificity via resolving JMJD2A complexed with mono-, di- and trimethyl forms of H3K9 and trimethyl form of H3K36[2]. Right after that, Couture at al reported the crystal structure of JMJD2A catalytic-core domain in complex with mono-, di- and trimethylated forms of H3K9 peptide[3]. In their work they demonstrated that JMJD2A recognized its substrates through a network of backbone hydrogen bonds and hydrophobic interactions that keeps the trimethyllysine into the active site.

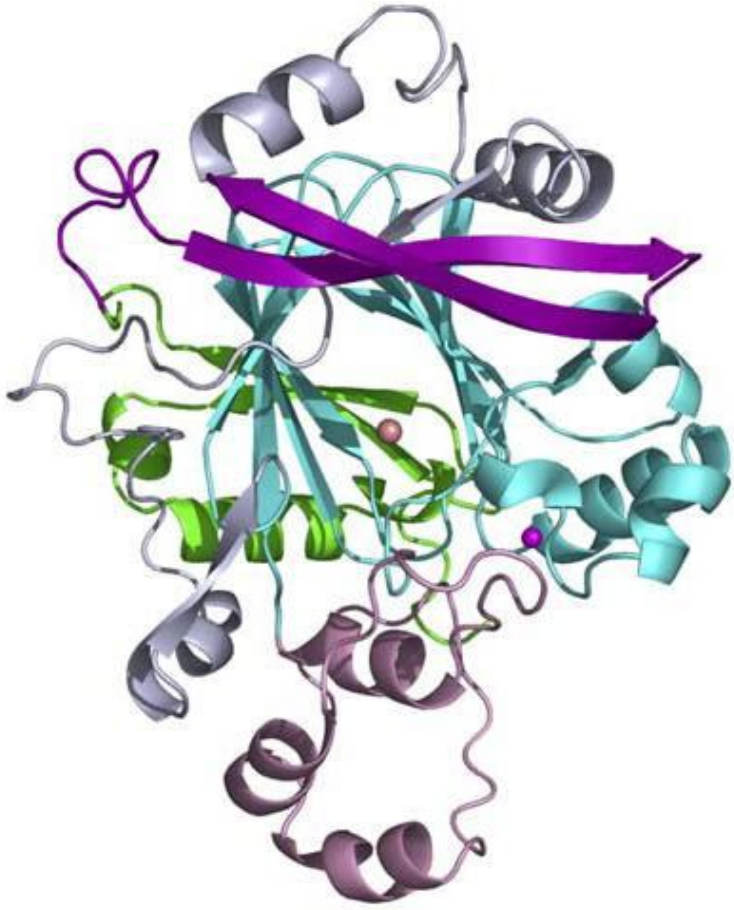

Figure 2.3: The catalytic-core domain of JMJD2A[10]. The domains include the JmjN domain (green), the long b hairpin (red), the mixed structural motif (gray), the JmjC domain (light blue), and the C-terminal domain (pink). The Fe(II) and Zn(II) ions are colored brown and purple, respectively.

2.2.2. Structural Basis of Substrate Binding

JMJD2A is a H3 trimethylated Lysine 9 (H3K9(me3)) and Lysine 36 (H3K9(me3)) specific demethylase[10]. Comparison between complex structures of the two substrates reveals that both peptides adopt bending β-strands but do not join β-sheet interaction with

the enzyme. Recognition of the substrates is obtained via a hydrogen bonds network and van der Waals contacts between the enzyme and the substrates. For both complexes the N-terminal residues of the substrates participate in van der Waals interactions with side chains of Ile 168 and the Val 313 of JMJD2A and the C-terminal residues of each peptide interact with Asn86, His240, Lys241 and Met242 in the enzyme. By contrast with van der Waals interactions, the hydrogen bonds network that anchors the peptides in substrate binding site is quite distinct. For the case of H3K9(me3) the enzyme binds to the peptide via hydrogen-bonding including main chain atoms of Gly33, Gly34, Val35, Lys36(me3), Lys37 and His39, whereas backbone hydrogen-bonding is established via the main chain atoms of trimethyl-Lys9, Thr11, Gly12 and Gly13 in the H3K9(me3) case [3].

JMJD2A makes few contacts with the side chain atoms of the H3 trimethyl-K36 and H3 trimethyl-K9. In the case of H3K36(me3) Lys37 of the peptide is within van der Waals contact range of Ala 134 and Asp 135. Similarly few interactions are determined between JMJD2A and the side chains of the residues neighboring H3 trimethyl-K9 site [3]. However guanidinium group of Arg8 is found within hydrogen-bonding distance of carboxyl group of Glu169 of the enzyme, Couture at al reported alternative conformations for side chain of this residue[3].

It was revealed by both Couture at al. and Ng at al that JMJD2A shows maximal activity with the trimethylated H3K9 peptide substrate, suggesting that this sequence adopts an optimal conformation [2-3]. The H3K9(me3) substrate gains a broad ‘W’-shaped conformation during binding. This bent peptide conformation is stabilized by intra-substrate hydrogen bonds and required for sequence specificity of JMJD2A [2]

2.2.3 Methylation-state Specificity of JMJD2A

It was reported by Whetstine at al. that JMJD2A is Lys(me3) specific enzyme and only had activity on Lys(me2) in the presence of excessive amount of enzyme and had no activity on Lys(me1)[7]. Thereupon methylation-state specificity of JMJD2A became a

object of interest for different research groups. Couture at al resolved crystal structure of JMJD2A in complex with H3 mono-, di-, trimethyllysine 9 and H3 trimethyllysine 36[3, 7]. By comparison of H3 trimethyllysine 9 and H3 trimethyllysine 36 complex structures they found out that the substrates were oriented within the binding-cleft through main chain hydrogen bonds with the Glu169 carboxyl oxygen and Tyr175 hydroxyl group respectively in the enzyme. The aliphatic groups of trimethyllysine side chain were found in contact with Asp191, Lys241 and Asn290 in JMJD2A and the trimethylammonium head is settled in an oxygen-encircled cavity next to the Fe(II) center. The oxygen-encircled cavity which was referred by Couture at al. as methylammonium binding pocket was composed of the carbonyl oxygen of Gly170, the hydroxyl groups of Tyr177 and Ser288 and the carboxylate side chain of Glu190[3]. They demonstrated the interactions between methyl groups of H3 trimethyllysine 9 and the oxygens lining the methylamonium pocket and defined these interactions as CH---O hydrogen bonding[4]. They emphasized the importance of these hydrogen bonds in methylation-state specificity via comparing H3 trimethyl and dimethyl forms of complexes.

Ng and coworkers argued the importance of the water molecules involving in the orientation of the substrate methyl groups in the catalytic pocket for methylation-state specificity of JMJD2A[2]. In their work, they demonstrated that at the case of H3K9(me1) and H3K9(me2) the niche of absent methyl groups were occupied by water molecules and related these findings to substrate specificity of JMJD2A.

Chapter 3

COMPUTATIONAL METHODS

3.1. System Setup

3.1.1 Preparation of Initial Coordinate Files

The initial coordinates of MD simulations were taken from X-ray structures of complexes between JMJD2A and H3 tail that mono-, di-, trimethylated at K 9 corresponding to entries 2OT7 at 2.13 Å, 2OX0 at 1.95 Å and 2OQ6 at 2 Å respectively in Protein Data Bank (PDB). Although each initial structure was containing two copies of complex, based on a previous work[10] that reports JMJD2A functions as monomer we extracted the B chain of JMJD2A and conjugate H3 peptide. Due to all B chains and conjugate peptides have the same amino acid types and number we did not further process the complexes. All histidine residues that do not contribute to coordination of Fe(II) set to neutral and protonated at N_ε . The protonation state of His 240 and His 188 that are coordinating the cation were determined considering a previous publication[12] which states that all mononuclear Fe ions adopt N_ε tautomeric conformation of histidines. All crystallographically resolved water molecules were retained in the systems. Each system was then solved using TIP3P[13] water in a cubic box with at least 10 A distances around the solute. Each composite system contains approximately 45,000 atoms. The Amber03[14] force field was used with tLeap in Amber10[15] for system set up. For free energy calculations the ligands were extracted from complexes and simulated separately.

3.1.2. Parameterization of Non-standard Residues

In our systems we have three non-standard residues (monomethylated lysine, dimethylated lysine and trimethylated lysine), which are modified forms of lysine and the cofactor of the enzyme JMJD2A, N-oxalylglycine (NOG). To perform MD simulations, together with the partial charges of the non-standard residues and the cofactor the force constant of bonds, angles and dihedrals of the cofactor need to be defined.

For consistency, to derive the atomic charges of non-standard residues we obeyed the protocol proposed by Duan at al[14]. In these calculations each amino acid was characterized by a dipeptide fragment, which comprised the amino acid, N-methyl (Nme) group at N-terminal and acetyl group (Ace) at C-terminal (see Figure I). For each dipeptide, two conformations with main-chains dihedral angles (phi, psi) constrained to alpha helical and beta stranded conformations were prepared and subjected to energy minimization by using the program Accelrys Discovery Studio[16]. These dipeptide conformers were then further geometry optimized at the HF/6-31G level of Quantum Mechanics (QM) Theory. All QM calculations were done by using the Gaussian simulation package 03[17]. Molecular electrostatic potential for each conformer was calculated by using density functional theory (DFT) method B3LYP with the ccpVTZ basis set. The IEFPCM continuum solvent model was employed to imitate an organic solvent environment (eps =4). Atomic charges were obtained by fitting the molecular electrostatic potential of dipeptides using the method Restraint Electrostatic Potential (RESP)[18]. The fitting phase was consisted of two stages. At the first stage, the conformers of each dipeptide were combined. At the second stage, the atomic charges of equivalent atoms were equated and the charges of terminal groups and heavy atoms were fixed. All fitting calculations were done with the RESP module of Amber10 suit of package. For preparation of Gaussian03 and RESP input files and processing of output files of these programs, the

version III of Resp Esp Charge Drive (R.E.D.)[19] program was used to accelerate the process.

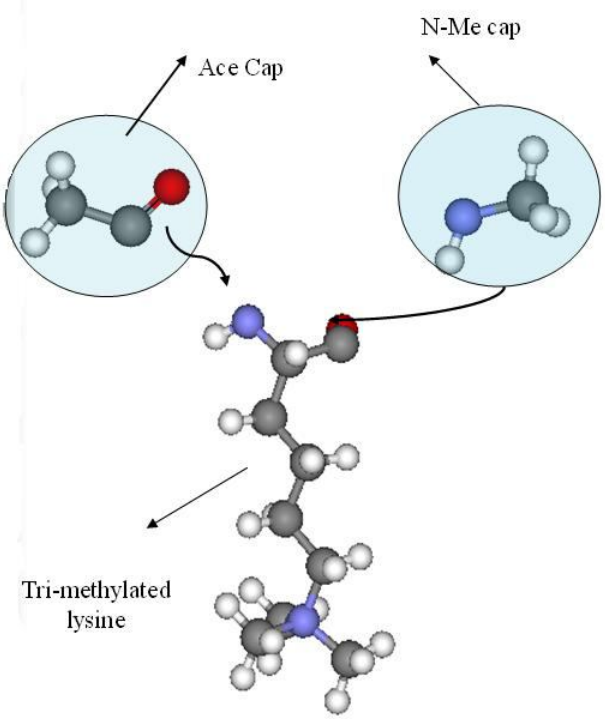

Figure 3.1: Preparation of trimethylated lysine for QM calculations.

For charge derivation of cofactor the same protocol with minor changes was used. A single conformation of NOG was prepared for QM calculations. Since the cofactor was a standalone molecule no blocking groups were added to the terminals.

Equilibrium values of the bond lengths, the angles and the dihedrals of the cofactor were taken directly from the optimized structure. The force constants of missing parameters were adopted from General Amber Force Field (GAFF)[20] using the analogy-based utility 'parmchk' of Amber 10 package.

3.1.3 Parameterization of Metal Centers

JMJD2A histone demethylase has a catalytic motif that includes Fe(II) and a zinc finger motif. In catalytic site Fe(II) is coordinated by two histidine residues (HIS 188 and HIS 276), one glutamic acid (GLU 190), one water molecule and the cofactor NOG (See Figure 3.2), whereas in coordination of Zn(II) three cystein residues (CYS 234, CYS 306 and CYS 308) and one histidine residue (HIS 240) have role[10]. Both Fe (II) and Zn (II) belong to the transition metals which are strongly interacting with surrounding molecules and causing large electronic rearrangements[21]. For this reason, to reproduce the parameters of metal centers different schemes which consider the charge transfer between metal and its coordination sphere were proposed:

i. Quantum mechanical description models
ii. Bonded models
iii. Non-bonded models

All these three models have some advantages and disadvantages. For our purpose the most convenient and applicable one was the non-bonded model, which proposed by Dal Pararo at al [21]. We applied this model for our case with some minor changes.

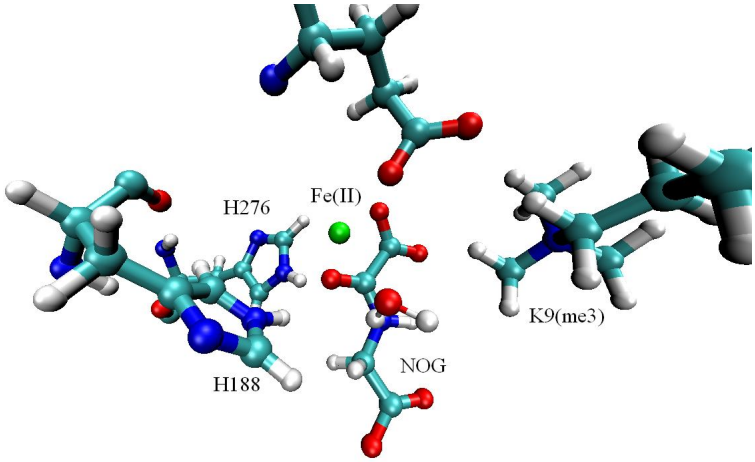

Figure 3.2: Fe(II) center.

For parameterization of metal centers, to avoid any bias the initial structure was taken from the PDB entry 2OQ7 (2.13 Å) which is not one of our initial structure of MD simulations. In the PDB file Ni(II) ion was replaced by a Fe(II) ion whereas the cofactor NOG was kept. Normally at physiological conditions the α-ketogluterate (α-KG) is found as cofactor. It was reported by a previous work that NOG occupies the same position and forms the same interactions with α-KG [1]. After the preparation phase the initial structure was subjected to a two stage minimization process by keeping the backbone atoms fixed. For minimization purpose Sander module of Amber 10 was used. Afterward the metal ions with their first coordination shell were extracted from the energy minimized structure. Zn(II) model was prepared for quantum mechanical calculations by cutting the amino acids from their C_β atoms. For Fe(II) center, together with the side chains of the amino acids, interacting portion of cofactor was also used. For geometry optimization the density functional theory method B3LYP with the base 6-31G** was used, constraining the C_β positions to mimic the protein scaffold. After structure relaxation, a single point calculation was performed using the B3LYP method and IEFPCM continuum solvent model at dielectric constant 4. This step was done using ccpVTZ basis set on all atoms except the metals, for which TVPZ basis set that is commonly used for transition metals was employed. Finally RESP method was used to obtain atomic point charges which reproduce the electrostatic potential calculated at the previous step.

3.2 Molecular Dynamics Concepts and Algorithms

Molecular dynamics (MD) is a sort of computer simulation in which atoms and molecules are allowed to interact for a period of time by approximations of known physics. It represents an interface between laboratory experiments and theory, and can be understood as a "virtual experiment".

MD inquires the relationship between molecular structure, movement and function.

The molecular dynamics simulation method is based on Newton's second law known as equation of motion, $F=ma$, where F is the force applied on the particle, m is its mass and a is its acceleration. If the force on each atom is known, it is possible to determine the acceleration of each atom in the system. Integration of the equations of motion then results in a trajectory that describes the positions, velocities and accelerations of the particles as they vary with time. The method is deterministic; once the positions and velocities of each atom are known, the state of the system can be predicted at any time.

Newton's equation of motion is given by;

$$F_i = m_i a_i \qquad \textbf{(3.1)}$$

where F_i is the force exerted on particle i, m_i is the mass of particle i and a_i is the acceleration of particle i. The force also can be expressed as the gradient of the potential energy.

$$F_i = -\nabla_i V \qquad \textbf{(3.2)}$$

Combining this to equations yields;

$$-\frac{dV}{dr_i} = m_i \frac{d^2 r_i}{dt^2} \qquad \textbf{(3.3)}$$

Where V is the potential energy of the system and r_i is the position of atom i. Newton's equation of motion can then relate the derivative of the potential energy to the changes in position as a function of time.

Therefore, to calculate a trajectory, one only needs the initial positions of the atoms, an initial distribution of velocities and the acceleration, which is determined by the gradient of the potential energy function. The equations of motion are deterministic; the positions and the velocities at time zero determine the positions and velocities at all other times, t. The

initial positions can be obtained from experimental structures, such as the x-ray crystal structure of the protein or the solution structure determined by NMR spectroscopy. The initial distribution of velocities are usually determined from a random distribution with the magnitudes conforming to the required temperature and corrected so there is no overall momentum;

$$P = \sum_{i=1}^{N} m_i v_i = 0 \qquad \textbf{(3.4)}$$

The velocities v_i, are often chosen randomly from a Maxwell-Boltzmann or Gaussian distribution at a given temperature, which gives the probability that an atom i has a velocity v_x in the x direction at a temperature T.

$$P(v_{ix}) = \left(\frac{m_i}{2\pi k_\beta T} \right)^{1/2} \exp\left[-\frac{1}{2} \frac{m_i v_{ix}^2}{2 k_\beta T} \right] \qquad \textbf{(3.5)}$$

3.2.1. Force Field-AMBER 03

In this work we used AMBER 03 force field which is an effective two-body model. This force field like other commonly used classical force fields, consists of two parts; the functional form and the parameters. The functional form is the description of the potential function. The potential function describing interactions among particles comprises electrostatic, van der Waals, bond, bond angle, and dihedral terms and can be shown as:

$$E_{total} = \sum_{bonds} K_b (b - b_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}} \right] \quad \textbf{(3. 6)}$$

where, K_b and K_θ are the force constants for the bond and bond angles, respectively; b and θ are bond length and bond angle; b_{eq} and θ_{eq} are the equilibrium bond length and bond angle; Φ is the dihedral angle and Vn is the corresponding force constant; the phase angle takes values of either 0° or 180°. The non-bonded part of the potential is represented by van der Waals (Aij) and London dispersion terms (Bij) and interactions between partial atomic charges (q_i and q_j). ε is the dielectric constant that takes into account of the effect of the medium that is not explicitly represented[14].

In respect of parameters, the atomic partial charges were obtained by fitting to the molecular electrostatic potentials of dipeptides calculated using B3LYP/cc-pVTZ//HF/6-31G** level of theory. The main-chain torsion parameters were obtained by fitting to the energy profiles of Ace-Ala-Nme and Ace-Gly-Nme dipeptides calculated using MP2/ccpVTZ// HF/6-31G** quantum mechanical methods. All other parameters were taken from the existing AMBER force fields. The major difference from previous force fields is that all quantum mechanical calculations were done in the condensed phase with continuum solvent models and a dielectric constant of $\varepsilon = 4$.

3.2.2. Full Electrostatic Computation

In our simulation Columbic interactions were treated with the particle mesh Ewald (PME)[22] method. In PME method the interaction potential is separated into two terms;

$$\Phi(r) = \Phi_{sr}(r) + \Phi_{lr}(r) \quad \textbf{(3.7)}$$

A short-ranged part $\Phi_{sr}(r)$ that sums quickly in real space and and a long-ranged part $\Phi_{lr}(r)$ that sums quickly in Fourier space. In equation (3.7) r stands for position. The particle mesh Ewald method is based on the idea that the direct summation of interaction energies between point particles;

$$E_{TOT} = \sum_{i,j} \varphi(r_j - r_i) = E_{sr} + E_{lr} \qquad \textbf{(3.8)}$$

can be replaced with two summations; a direct sum E_{sr} of the short-ranged potential in real space;

$$E_{sr} = \sum_{i,j} \varphi_{sr}(r_j - r_i) \qquad \textbf{(3.9)}$$

and a summation in Fourier space of the long ranged-part;

$$E_{lr} = \sum_{k} \hat{\Phi}_{lr}(k) \left| \hat{\rho}(k) \right|^2 \qquad \textbf{(3.10)}$$

where $\hat{\Phi}$ and $\hat{\rho}(k)$ represent the Fourier transform of the potential and the charge density respectively. In equations (3.8) and (3.9); r_i and r_j correspond to positions of particle i and j respectively. In order to evaluate the Fourier transform $\hat{\rho}(k)$ of the charge density field Fast Fourier transform which requires the evaluation of the density field on a discrete lattice in space is used.

3.2.3. Simulation Details

All simulations were performed with the molecular dynamics package NAMD 2.6 together with the Duan at al force-field [14]. Bonded parameters were taken from GAFF and atomic partial charges for non-standard molecules were derived using the RESP procedure as proposed by Duan at al [14].

The systems were energy minimized for 10,000 steps with 2.6 version of NAMD[23]. Each system was annealed from 10 to 310 K over a period of 60 ps. The systems then equilibrated at 310 K using Langevin thermostat with a coefficient of 5/ps in the isobaric-isothermal (NPT) ensemble for 2 ns[24]. Periodic boundary conditions and the hybrid Nose-Hoover Langevin piston method[25-26] were used to control pressure at 1 atm. After equilibration, dynamics were continued with the same conditions that were used for equilibration for additional 18 ns. All hydrogen bond lengths were constrained with the SETTLE[27] algorithm. A multiple time-stepping algorithm was used, where bonded and the short-range non-bonded interactions were evaluated at every time step and the long-range electrostatic interaction were evaluated at every 2 time steps. A value of 2 ps was used for time step. In order to efficiently treat the electrostatic interactions PME was employed. To treat the short-range interactions a spherical cutoff of 12 Å was used.

3.3 Binding Free Energy-The MM/PBSA Approach

MM-PBSA[28] is a commonly used method to calculate binding free energy. For this method one needs dynamical sampling of the complex system and post processing of the trajectory structures. The binding free energy may be calculated by comparison of complex trajectory with separate trajectories of receptor and ligand or more frequently from a single trajectory of complex. In our study we used the single trajectory method.

The binding free energy of two molecules (receptor and ligands) is the difference between the free energy of the complex and that of the receptor and the ligand which can be shown as;

$$\Delta G = \overline{G}_{complex} - \overline{G}_{receptor} - \overline{G}_{ligand} \quad \textbf{(3.11)}$$

According to this method the free energy of each species is calculated as follows

$$\overline{G}_{molecule} = \overline{E}_{MM} + \overline{G}_{PBSA} - TS_{MM} \quad \textbf{(3.12)}$$

Here $\overline{G}_{molecule}$ and $\overline{E}_{MM}$ denote the computed average free energy and average molecular mechanical energy,

$$\overline{E}_{MM} = \overline{E}_{bond} + \overline{E}_{angle} + \overline{E}_{tors} + \overline{E}_{vdw} + \overline{E}_{elec} \quad \textbf{(3.13)}$$

where above elements correspond to bond, angle, torsion, van der Waals and electrostatic terms in the molecular mechanical force field. $\overline{G}_{PBSA}$ is the solvation free energy. This term has two parts.

$$\overline{G}_{PBSA} = G_{elect} + G_{non-elect} \quad \textbf{(3.14)}$$

The First part (G_{elect}) is the polar part and comes from a numerical solution of Poisson-Boltzmann equation. The second part ($G_{non-elect}$) is the cost of opening a cavity in the condensed phase and generally is estimated as product of the surface area and an effective surface tension term[29].

$$\overline{G}_{non-elect} = \gamma SASA + b \quad \textbf{(3.15)}$$

In equation (3.15) γ is surface tension parameter set to 0.0072, *SASA* is the solvent accessible surface area of molecule of interest determined by ICOSA[30] method and *b* is a parameterized value, set to 0.0 for both PB and GB in this study.

3.4. Entropy Calculations

To understand the energetic determinants of substrate recognition at molecular level we need to consider the enthalpic and entropic component of binding. Enthalpic contributions provide a measure of the strength of interactions between the receptor and the ligand. Entropic contributions involve the change in solvent entropy and the loss of solute conformational degrees of freedom.

Equation (3.12) can be written as;

$$\overline{G}_{molecule} = H_{molecule} - TS_{MM} \qquad \textbf{(3.16)}$$

Where $H_{molecule}$ and TS_{MM} are the enthalpic and the entropic contributions respectively. The entropy term S_{MM} comprises translational, rotational, vibrational and configurational entropy of solute and can be expressed as;

$$S_{MM} = S_{trans} + S_{rot} + S_{vib} + S_{config} \qquad \textbf{(3.17)}$$

S_{trans} and S_{rot} are the entropic contributions from translational and rotational motions, respectively. In this study these quantities were calculated from their gas phase partition function, Q, according to following classical statistical mechanics relations.

$$S_0 = k \ln Q - \left(\frac{\partial LnQ}{\partial k_B T} \right)_v \qquad \textbf{(3.18)}$$

where

$$Q_{trans} = \frac{V}{h(K_B T / 2\pi m)^{3/2}} \qquad \textbf{(3.19)}$$

$$Q_{rot} = \left(\frac{\pi^{1/2}}{\sigma}\right)\left(\frac{1}{hck_BT}\right)^{3/2}\left(\frac{1}{I_AI_BI_C}\right)^{1/2} \quad \textbf{(3.20)}$$

In above equations, V is the volume, h is Planck's constant, k_B is the Boltzmann constant, T is temperature, m is the molecular mass, σ is the symmetry number, c is the speed of light, and I_x are the three inertial rotational constants.

S_{vib} is the entropy contribution of vibrational motion. In this study we utilized normal mode analysis to obtain the vibrational entropy.

$$S_{vib} = \sum_{i=1}^{3N-6}\left[\frac{h\nu_i}{k_BT(e^{\nu_i/k_BT}-1)} - \ln(1-e^{-\nu_i/k_BT})\right] \quad \textbf{(3.21)}$$

S_{config} is the configurational entropy from the side chains reorganization effects which was ignored in this study.

Due to the high computational demand, entropy calculations were performed only for a few snapshots which may cause a sampling problem. In order to overcome this problem we followed a new approach. In this approach firstly we clustered the snapshots in 5 classes according to their similarities and provide a representative structure for each class[31]. Then we minimized each representative structure in gas phase using the conjugate gradient method until the root-mean-square of the elements of the gradient vector is less than 10^{-3} kcal $mol^{-1}A^{-1}$. For each minimized snapshot the entropy contributions of each component is computed at 300 K using the NMODE module of Amber8[32]. After all, the entropy contribution of each cluster to final entropy was obtained via weighting the relevant entropy value with the occurrence of the cluster.

3.5. Computational Alanine Scanning

Computational alanine scanning method[33] is a widely used approach to estimate the hot spots of protein-protein interfaces. In this approach the amino acid of interest is mutated to alanine and the binding free energy difference between the wild type and mutated one is calculated according to this equation:

$$\Delta\Delta G_{binding} = \Delta G_{binding-mutant} - \Delta G_{binding-wildtype} \quad \textbf{(3.22)}$$

where $\Delta\Delta G_{binding}$ is the difference between binding free energy, $\Delta G_{binding-mutant}$ and $\Delta G_{binding-wildtype}$ are binding free energies of mutant and wild type respectively.

In our study we performed the computational alanine scanning mutagenesis in two ways. For the first and simplest way we extracted the mutated structures from the structures of wild type via truncation of side chains and calculated the binding free energy. For the second way we used the initial structure of wild type to provide a starting structure for the mutant and performed MD simulation. Then we extracted the structures from the resulting trajectory. For both approaches we calculated the enthalpic term of binding free energy via MM-PBSA method that we already mentioned.

3.6. Binding Free Energy Decomposition

Decomposing ΔG in terms of contributions from residues of both binding partners provides insight into origin of binding on an atomic level. Based on this, Golkhe and coworkers[34] proposed a scheme which enables us to investigate the contribution of each residue to the binding free energy by means of component analysis. Reconsidering equation (3.11) in this respect we reproduce:

$$\Delta G = \sum_{\substack{j\in complex\wedge \\ j\in receptor}} (\langle G_{complex}(i,j)\rangle - \langle G_{receptor}(i,j)\rangle)$$
$$+ \sum_{\substack{j\in complex\wedge \\ j\in ligand}} (\langle G_{complex}(i,j)\rangle - \langle G_{ligand}(i,j)\rangle) \qquad (3.23)$$
$$- H_{trans/rot}$$

Here, j stand for residues either located on receptor or protein and i stand for number of snapshots. There by each addend provide the contribution to binding free energy by residue j averaged over snapshot i. Since the energy contribution due to translational and rotational degrees of freedom is already considered in above equation, $G_{molecule}(i,j)$ includes contributions from internal gas-phase energy, solvation free energies and entropies:

$$G_{molecule}(i,j) = H^{gas}_{molecule}(i,j) + G^{solvation}_{molecule}(i,j) - TS_{molecule}(i,j)$$

Considering the fact that the total electrostatic work of creating a given charge distribution within a solute, which buried in a solvent is a quadratic function of charges in the GB formula, a GB analogue of the electrostatic potential can be defined at the r_l position of each atom:

$$\Phi(r_l) = -\sum_k (1 - \frac{\exp(\kappa f^{GB}_{kl})}{\varepsilon_w}) \frac{q_k}{f^{GB}_{kl}(r_{kl})} + \sum_{k\neq l} \frac{q_k}{r_{kl}} \qquad (3.24)$$

where ε_w is the dielectric constant of the solvent and q_k are atomic partial charges and κ is Debye-Hückel screening parameter. In equation (3.24) f^{GB}_{kl} is a certain smooth function which is assumed to depend only upon atomic radii and interatomic distances r_{kl} [35]. Thus the electrostatic contribution from residue j is obtained according to:

$$G_{elec}^{molecule}(i,j) = -\frac{1}{2}\sum_{l\in j}\sum_{k}(1-\frac{\exp(-k)}{\varepsilon_w})\frac{q_k q_l}{f_{kl}^{GB}(r_{kl})} + \frac{1}{2}\sum_{l\in j}\sum_{k\neq l}\frac{q_k q_l}{r_{kl}} \quad \textbf{(3.25)}$$

The first term on the right hand side of above equation determines the electrostatic contribution to the free energy of solvation for residue j of snapshot i; whereas the second term stands for the gas-phase charge–charge interactions.

The solvent-accessible surface area per atom was computed with a recursive algorithm[30] that uses icosahedra centered on this respective atom as starting point. In every step each triangular face of the polyhedron is divided into four pieces of equal size, thus a better approximation of a sphere is obtained.

Chapter 4

RESULTS AND DISCUSSION

4.1 Structural Analyses

In order to gain insight into how JMJD2A discriminates between its substrates we analyzed positional fluctuations, located the hydrogen bonds and calculated some critical distances. During 18 ns of MD simulation, all three enzyme-substrate trajectories exhibit backbone root-mean-square deviation (rmsd) values below 2 Å which indicates stability (See Figure 4.1). Backbone fluctuations show minor changes in the last 6 ns and 2ns of simulations for H3K9(me3) and H3K9(me2) respectively.

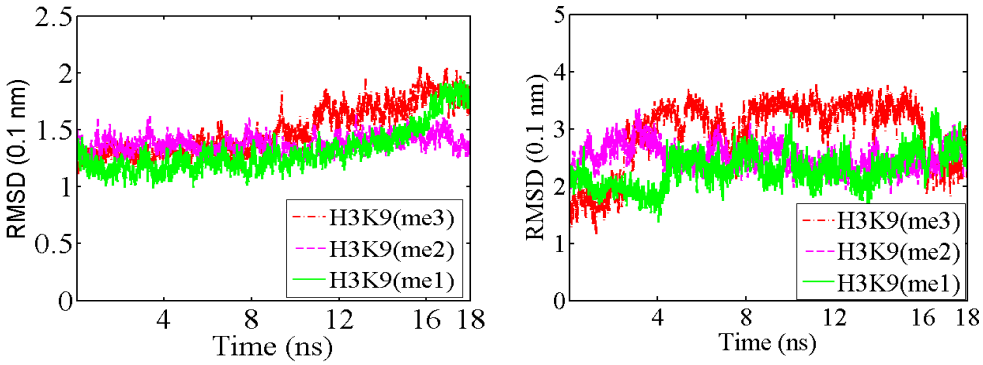

Figure 4. 1 Left Panel: Backbone RMSD of the Enzyme-substrate complexes during the MD production stage. Right Panel: Backbone RMSD of the bonded substrates throughout the MD simulations. For clarity the data have been plotted with a time interval of 4 ps.

To reveal the behavior of bonded substrates we also calculated the RMSD of these peptides and provided the results in right panel of Figure 4.1. As seen in this figure H3K9(me3) exhibit higher RMSD values compared to other two peptides.

4.1.1. Structure of Fe(II) Center

JMJD2A histone demethylase has a catalytic motif that includes Fe(II). In catalytic site Fe(II) is coordinated by two histidine residues (His188 and His276), one glutamic acid (Glu190), one water molecule and the cofactor NOG in a bidentate way (See Figure 4.2). In order to correctly represent the physics of Fe(II) center we parameterized the metal center considering the scheme proposed by Dal Pararo at al[21]. This scheme bases on charge transfer between metal and its coordinating environment. We analyzed the behavior of the metal center throughout the 18 ns of MD simulations for H3K9(me1), H3K9(me2) and H3K9(me3) systems. For each system we found out that the mean fluctuation of Fe(II) cation is less than 1 Å which demonstrates the success of the simple method. We further assessed the structure of the metal center taking H3K9(me3) as model. During the 18 ns of simulations the overall structure of the metal center remained almost the same (See Figure 4.2). It was interesting to look at the interactions between Fe(II) cation and its coordinating water molecule. This crystal water molecule together with equilibrium phase remained bonded to Fe(II) almost for 16 ns which is indicative of a strong interaction. Before dissociation of the crystal water molecule, a second water molecule came and bond to Fe(II). For a while the two water molecules coordinated Fe (II) and thereafter the crystal water molecule broke away.

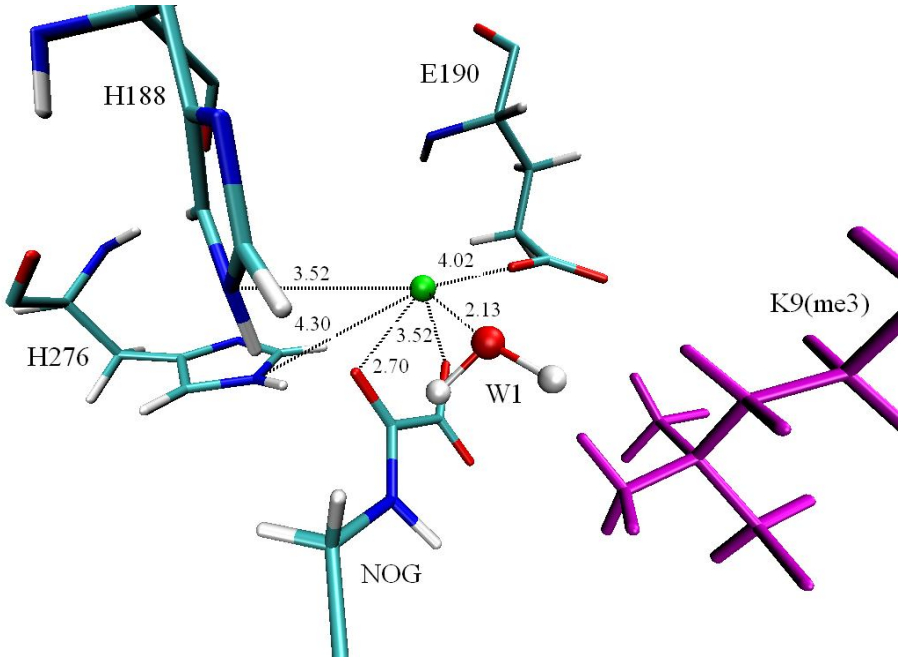

Figure 4. 2 Coordination of Fe(II) in active site. Fe(II) is penta-coordinated by NE(His188), NE(His276), OE1(Glu190), one water molecule(WAT1), O2(OGA) and O2'(NOG). Fe(II) cation is shown by light green sphere, trimethylated Lys9 side chain are showed by magenta. Other amino acids and the cofactor NOG are shown by licorice representation with the atom type color code (O: Red, N: Blue, C: Green and H: White).

We also examined the behavior of Fe(II)-coordinating water molecule for H3K9(me1) and H3K9(me2) cases and provided the results in Figure 4.3. As seen in this figure for both cases the crystal water molecules remain bonded to Fe(II) throughout the simulations. For H3K9(me2) and H3K9(me1) the means of the distance between Fe(II) and the relevant water molecules are 2.15 (0,09) Å and 2,16 (0,10) Å respectively. For our satisfaction we probed the interaction between the Fe(II)-coordinating water and its surrounding for each case. Contrary to our expectation we did not detect any notable hydrogen bond between the water molecule and its surrounding atoms. It seems that van der Waals and Columbic

interactions between the water and Fe(II) cation anchor the water molecule in close proximity of Fe(II). For the three cases the Fe(II)-coordinating water molecule stands somewhere between Fe(II) and the methylammonium head. This finding is consistent with the mechanism that proposed for Fe(II)/α-ketoglutarate-dependent hydroxylases[36] .The site in which the Fe(II)-coordinating water molecule sits is most probably where the oxygen molecule binds and the activation starts[2].

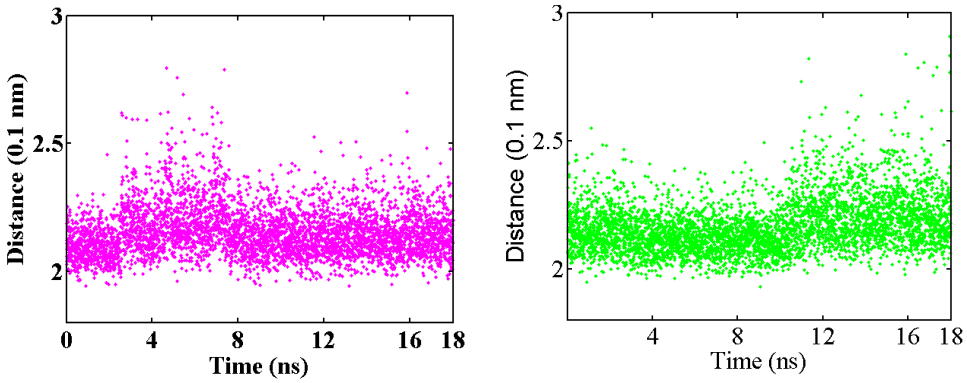

Figure 4. 3 The distance of Fe(II)-coordinating water molecules to Fe(II). Left panel: The change in Fe(II)-Wat1 distance throughout 18 ns of MD simulation for H3K9(me2) case. Right panel: Fe(II)-Wat1 distance change during 18 ns of production simulation for H3K9(me1). For clarity the data have been plotted with a time interval of 4 ps.

4.1.2. Role of Water Molecules

Ng and coworkers argued the importance of the water molecules that they detected in certain positions in crystallographic structure [2]. In their work, they demonstrated that at the case of H3K9(me1) and H3K9(me2) the niche of absent methyl groups were occupied by water molecules and related these findings to substrate specificity of JMJD2A. It remains unclear whether these water molecules are positionally stable and required for the discrimination of different methyl groups. We examined the behavior of these water molecules and their possible effects on the orientation of methyl groups throughout the

simulations. For the case of H3K9(me_1) two water molecules (W2, W3) in certain position were reported and the presence of these water molecules was associated with directing away the single methyl group from active site[2]. Wat3 remained around its initial position during the first half of the simulation. Within this time course W3 was in hydrogen bond interactions with Glu170 Tyr177 and Ser 288 (See Figure 4.4). Wat2 remained in close proximity of the methylammounium head throughout the simulation by forming hydrogen bond with Wat1 and Columbic interaction with Fe(II). To quantify the stability of Wat2 and Wat3 we computed the distance of those molecules to NZ of monomethyllysine and provided the results in A, B and C panels of Figure 4.5.

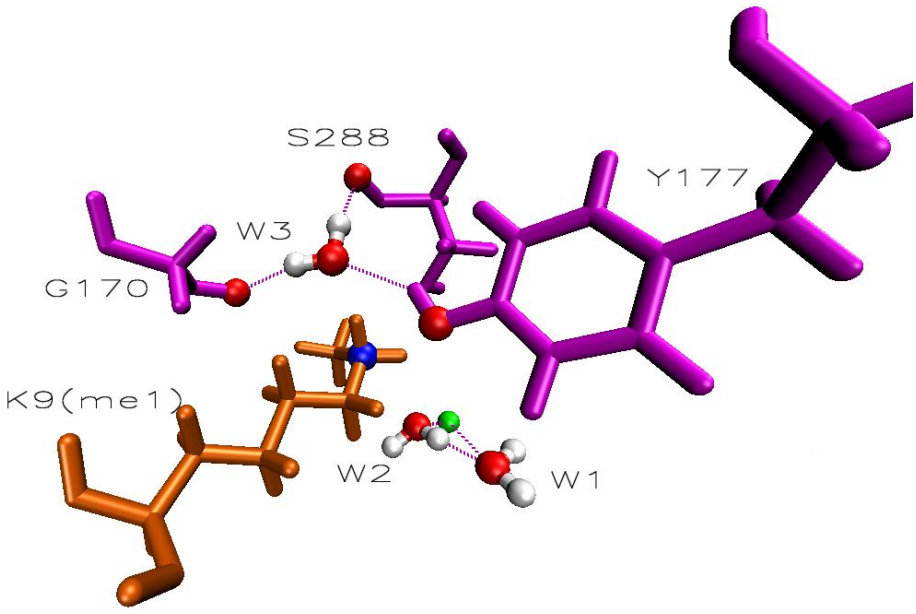

Figure 4.4 Representative snapshot of the water molecules in certain positions at H3K9(me1) case. Fe(II) cation is shown with green and critical oxygen atoms are shown with red spheres.

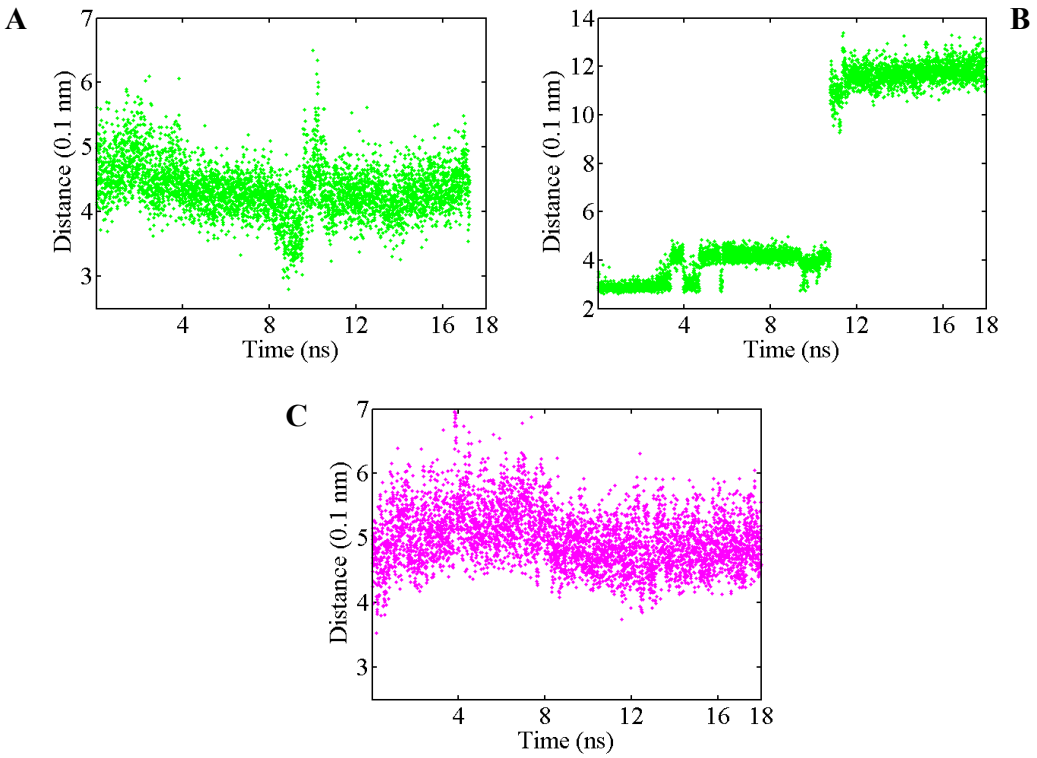

Figure 4. 5 The Distance of Critical Water Molecules to NZ(Lys9) for H3K9(Me_1) (upper panel A and panel B) and H3K9(me2) (panel C). For clarity a value of 4 ps is used for time interval. All distances are in Å

4.1.3 Methylammonium Binding Pocket

The NZ of the methyllysine locates in an oxygen-enclosed pocket that is formed by side chains of Tyr 175, Tyr177, Ser288, Asn290 and Glu190 and the backbone oxygen of Gly170. Those amino acids together provide appropriate orientation of the methyllysine side chain in the methylammonium binding pocket. The specificity of JMJD2A may stem from the interactions formed between methyllysine side chain and the oxygen rich

environment. We further assessed this issue and measured the distances between NZ and the surrounding oxygen atoms and pursued the possible hydrogen and C-H---O bonds for the three cases. In a previous study Couture at al. reported that the state specificity of JMJD2A results from the network of C-H---O hydrogen bonds in the methylammonium binding pocket[3]. Contrarily we did not detect any hydrogen bonds in this region. The bond distances were close enough while the angles were not appropriate to form C-H---O type hydrogen bonding[4, 37] despite the loose criteria we used. In Figure 4.6 we provided the mean distances between NZ and the surrounding oxygen atoms for each case. For the trimethyllysine case (See Figure 4.6 panel A) the oxygen atoms locate almost equally apart from NZ but for the dimethyllysine and monomethyllysine that is not the case. For the dimethyllysine case (See Figure 4.6 panel B) the distances between NZ and Gly170 and Asn290 oxygens become shorter compared to trimethylysine case. In comparison of monomethylysine and other two cases the decrease in the mean distance between NZ and Asn290 oxygen is prominent (See Figure 4.6 panel C) which is indicative of a strong interaction between these two groups. Moreover those distances ranges from 2.91 Å to 5.57 Å and are pretty enough to form Columbic interactions for the three cases. The importance of the Columbic interactions in state specificity of JMJD2A was emphasized in a previous study[1]. In this respect, our findings correlate with experimental results.

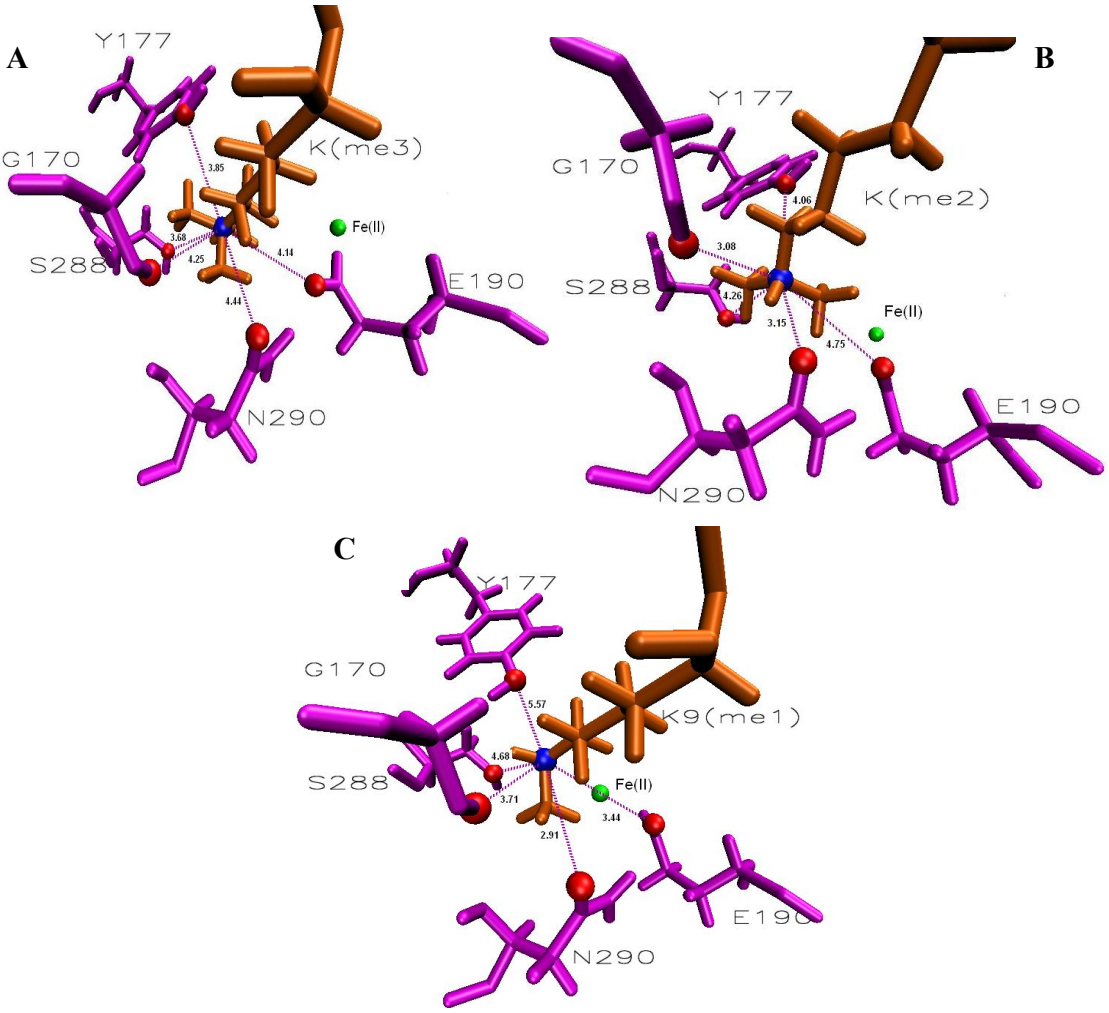

Figure 4.6 Methylammonium bindin pocket

To examine the behavior of methylammonium head in methylammonium binding pocket we measured the rotation around CE-NZ bond over time taking CD, CE, NZ and CZ1 as reference atoms for the three cases. Both the presence of water molecules and the variety of the critical distances between the methylammonium binding pockets might cause dissimilarities between cases. As seen in panel C of Figure 4.5 trimethyllysine head is almost free whereas dimethylysine (panel B of Figure 4.7) and monomethyllysine (panel A

of Figure 4.7) heads are almost completely restricted to rotate around CE-NZ bond. As can be seen in Figure 4.7 three discrete sites exist for NZ-methyl groups to occupy.

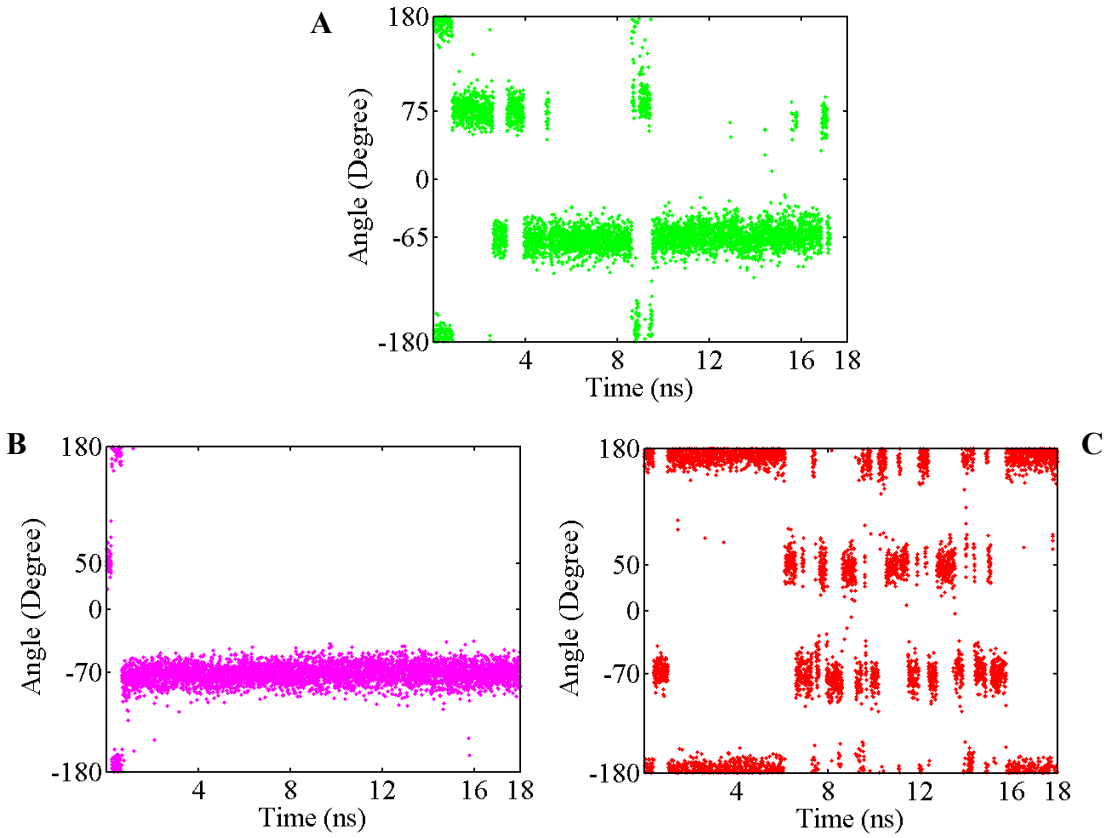

Figure 4. 7: Rotation around CE-NZ bond for H3K9(Me1) (panel A), H3K9(Me2) (panel B) and H3K9(Me3) (panel C). In order to measure the dihedral angle CD, CE, NZ and CZ1 atoms were used as reference atoms. For clarity a value of 4 ps is used for time interval. The angle values are in degree.

Reconsidering the results, which we obtained from the rotation around CE--NZ bond and the distance measurements in the methylammonium binding pocket we thought that the restriction on the angular motion in dimethyllysine and monomethyllysine cases could be an outcome of avoidance of steric clashes. Since methylammonium has symmetry around the NZ atom of methyllysine, 120^{o} rotation around the CE—NZ bond results in the same orientation. The equivalency of the occupied volume by each methyl group prevents the atoms from any steric overlaps during rotation. The value of CE-CD-NZ-X (X is the carbon atom of any methyl group) dihedral angle which allows the relevant methyl group to align with Fe(II) is 170^{o}. For the case of trimethyllysine always one methyl group provides the effectual value. The CE-CD-NZ-CZ1 dihedral angle was restricted to around 70^{o} and 65^{o} for di- and monomethylysine cases respectively. In order to further asses the space problem that arises from steric barriers we picked representative structures and turned the methylammonium head around the CE-NZ bond with -120^{o} each time and defined the steric clashes. As we expected, when we rotated the trimethylammonium head we did not detect any steric overlaps. However for the dimethyl and monomethyllysine cases we found out several overlaps. For the dimethyllysine case the -120^{o} artificial rotation of the methylammonium head around the bond caused the second methyl group atoms to overlap with Asn290(OD1) and Gly170(O) atoms (See Figure 4.8 A2 panel). Another -120^{o} artificial rotation created the similar overlaps with the first methyl group atoms (See Figure4.8 B2 panel).

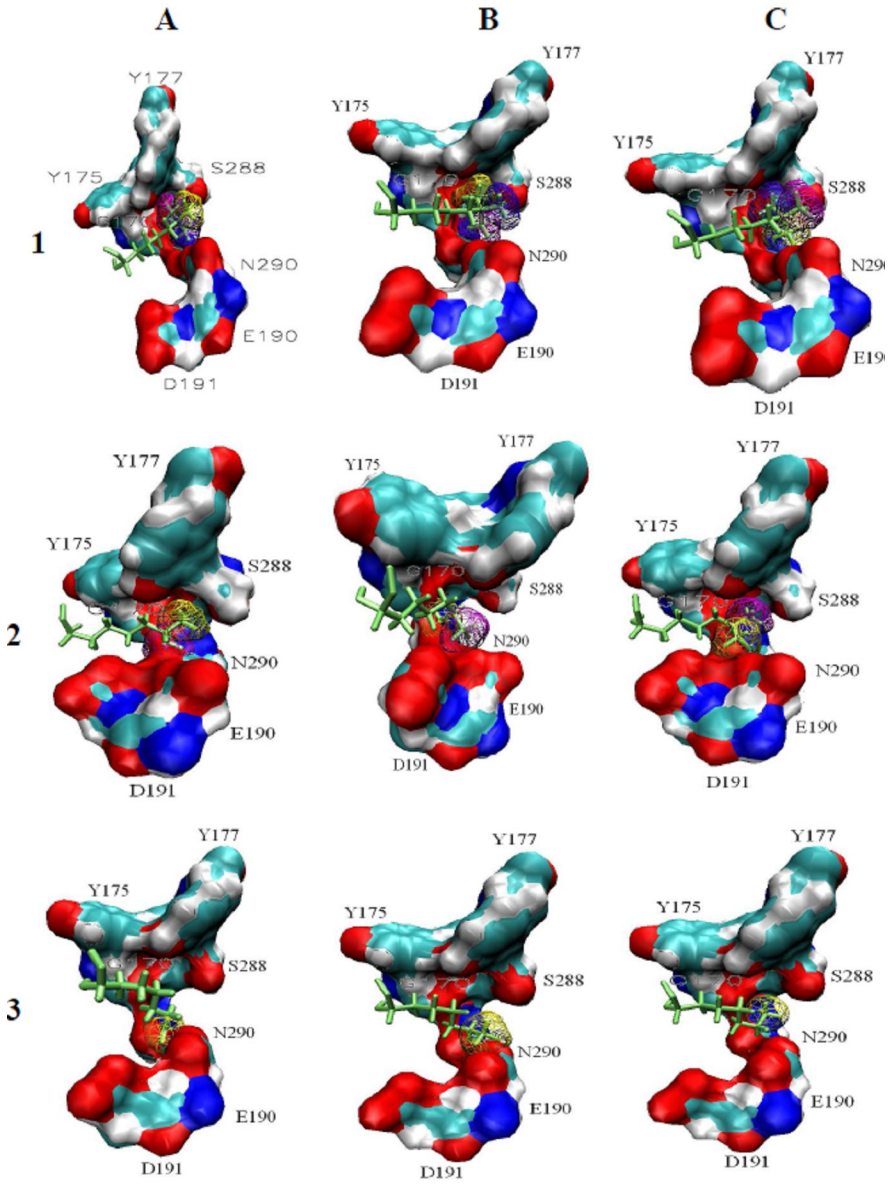

Figure 4.8 Column A: The orientation of methyllysine head at the CD-CE-NZ-CZ1 dihedral angle value 175°. Column B: CD-CE-NZ-CZ1 dihedral angle value 50°. Colum C: CD-CE-NZ-CZ1 dihedral angle value -70°. From up to down: H3K9(me3), H3K9(me2), H3K9(me1).

The -120^{o} rotation which set the dihedral angle value to 175^{o} resulted in overlaps of single methyl group atoms and Asn290(OD1,HD2) and Glu190(OE1) (See Figure 4.8 A3 panel). Another -120^{o} rotation around methylammonium head caused the Glu190(OE1,OE2) and Asn290(OD1) to overlap with the methyl group atoms (See Figure 4.8 B3 panel).

In order to handle the issue from another point of view, we determined certain states and from the occurrences of these states we obtained relative energy values by using the Boltzmann distribution equation. We computed the occurrence of each state by utilizing last 14 ns of the trajectory. We used this part of trajectories for two reasons. First the bound form of the trimethllysine substrate came to equilibrium at 4th ns of the production phase. (See Figure 4.1 right panel). Second this part of the trajectory makes the occurrence of each methyl group of the trimethyllysine equal which is parallel to our intuition. The states were obtained by discetizing the range (-180^{o})-(180^{o}) with an interval of 30^{o}. As seen in Figure 4.9 the trimethylysine head has three minima which separated from each other by approximately equal energy barriers. Since the energy barriers have equal values the probability of the transition from one state to another is equal. As we see in the same figure monomethyllysine head also has three minima but the energy barriers show variety. One minimum that coincides with -65 is quite lower compared to other two. For the case of dimethylysine one single minimum is seen and the energy barriers are very high compared to other two cases. Most probably the reason which impairs the symmetry between the minima for the case of di- and monomethyllysine is the steric overlaps that we already detected. During the rotation the side chains of the residues which cause overlaps with the methyl groups must not find a more favorable location. Displacements of the side chains increase the amount of required energy to rotate the methyl ammonium head.

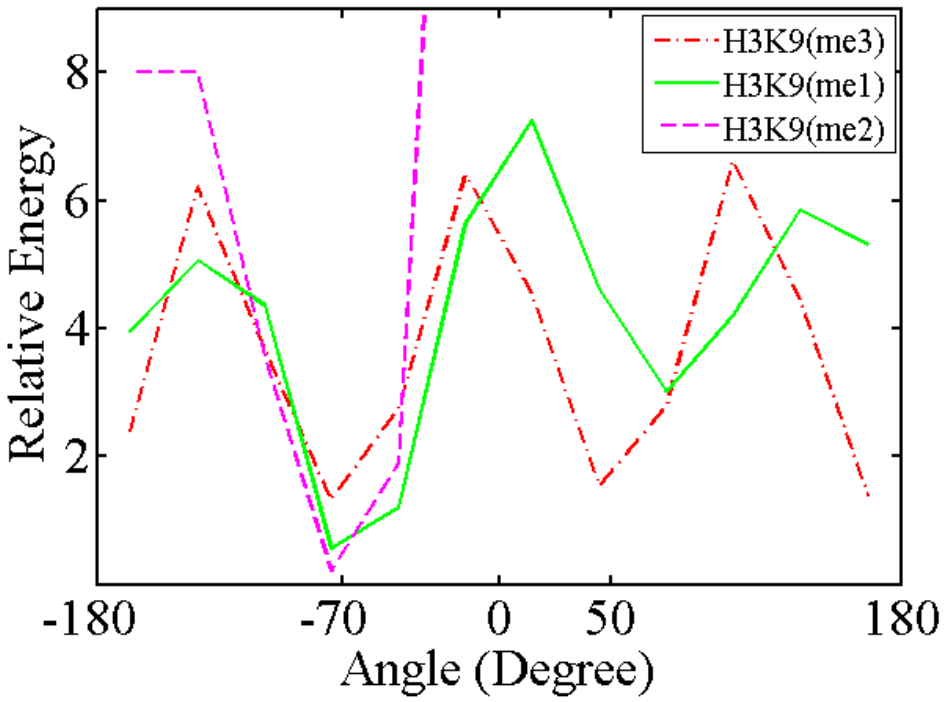

Figure 4.9 : Relative energy values of the three cases in terms of CD-CE-NZ-CZ1 angle. The range (-180°)-(180°) was discritized with an interval of 30° to define the states. From occurrence of each state by using the Boltzmann distribution equation, the relative energy of each state was calculated.

Besides the orientation of the methyl groups relative to Fe(II), the distance between the methyl groups and Fe(II) may be an explanation of activity specificity of JMJD2A. In order to assess this subject we determined a sphere surrounding Fe(II) and quantified the occurrence of each methyl carbon inside the sphere for the three cases. However we obtained same trend for different values of sphere radius, we provided the results for the value of 4.7 Å such that the occurrence of the methyl carbons of trimethyllysine approximately to be equal. We observed a significant difference between the H3K9(Me_3)

case and other two cases for finding a methyl carbon closer than 4.7 to Fe (II) (see Table 4.I).

Table 4.1 : The occurrence of methyl groups in a certain proximity of Fe(II). A sphere of radius 4.7 Å surrounding Fe(II) is determined and the occurrence of methyl carbons inside the sphere for three cases is calculated. All values are in percentage.

	CZ1(%)	CZ2(%)	CZ3(%)	TOTAL(%)
H3K9(Me3)	32,0	31,0	23,0	**86,0**
H3K9(Me_2)	4,5	0	-	**4,5**
H3K9(Me1)	1,5	-	-	**1,5**

The occurrence of methyl groups in a certain proximity of Fe(II) was in good agreement with presence of water molecules and the orientation of methyl groups. For H3K9(me3) both the presence of three methyl groups and the absence of water molecules which may restrict orientation of methyl groups, raise the occurrence of methyl groups in close proximity of Fe(II). Also the increase in atom numbers in methylammonium head might raise the van der Waals interaction strength between the head and the iron ion.

4.1.4. Hydrogen Bonds Analyses

It was reported by Chen at al[1] that the entire interactions between JMJD2A and its substrates involve 10 hydrogen bonds and one salt bridge. But those results come from analyses of crystal structures of trimethylated peptides in complex with JMJD2A. We questioned whether there is any difference between H3K9(me1) H3K9(me2) and H3K9(me3) in hydrogen bonding at enzyme-ligand interface. In order to select hydrogen bonds we used the criteria hydrogen-bond distance, $R< 3.00$ A and hydrogen-bond angle,

120 < θ < 180. We computed hydrogen bonds for first 6 ns of simulations for the three cases and provided the results in Table 4.2.

Table 4.2 : Hydrogen bonds involving JMJD2A, with respective occupancies, distances and deviations from linearity, in H3K9(me3) (red) H3K9(me2) (magenta) and H3K9(me1) complexes. Hydrogen bonds are defined by the donor-acceptor distance less than 3Å and donor-hydrogen-acceptor angle grater than 120°. Only bonds with occupancies higher than 20% are shown. The occupancies are computed from the first 6 ns of production simulations. The substrate residues are shown in bold.

SYSTEM	Donor	Acceptor	% Occupancy	Distance	Angle
H3K9(me3)	Glu169(O)	**Lys9(N)**	56,36	2.85 (0.08)	24.49 (12.49)
H3K9(me3)	Asp311(O)	**Arg8(N)**	51,70	2.87 (0.08)	29.53 (11.62)
H3K9(me3)	Asp135(OD2)	**Thr11(N)**	46,90	2.86 (0.08)	22.37 (12.47)
H3K9(me3)	**Gly13(O)**	Asn86(ND2)	38,37	2.88 (0.08)	22.14 (11.43)
H3K9(me3)	Asp135(OD1)	**Thr11(OG1)**	32,45	2.73 (0.13)	22.94 (12,41)
H3K9(me3)	**Lys14(O)**	Arg309(NH2)	27,58	2.78 (0.10)	29.22 (10.56)
H3K9(me3)	Lys9(O)	**Arg8(NE)**	23,32	2.87 (0.08)	26.64 (12.11)
H3K9(me3)	**Ser10(O)**	Lys241(NZ)	21,25	2.81 (0.09)	29.51 (13.82)
H3K9(me3)	**Gly14(O)**	Arg309(NH1)	20,45	2.81(0,09)	31.10 (14.66)
H3K9(me2)	Asp311(O)	**Arg8(N)**	60,36	2.87 (0.08)	27.24 (11.60)
H3K9(me2)	**Lys9(O)**	Arg8(NE)	57,43	2.86 (0.08)	27.03 (12.11)
H3K9(me2)	**Ala7(O)**	Glu169(N)	45,70	2,88 (0,08)	26,96 (11,04)
H3K9(me2)	**Gly13(O)**	Asn86(ND2)	38,37	2.88 (0.08)	22.14 (11.43)
H3K9(me2)	**Arg8(O)**	Met313(N)	35,90	2,80 (0,10)	35,61 (13,30)
H3K9(me2)	Glu169(O)	**Lys9(N)**	33,51	2.90 (0.07)	25.80 (11.96)

SYSTEM	Donor	Acceptor	% Occupancy	Distance	Angle
H3K9(me2)	Asp135(OD1)	**Thr11(OG1)**	32,51	2.75 (0.12)	16.24 (8,85)
H3K9(me2)	**Ser10(O)**	Lys241(NZ)	29,91	2,82 (0,09)	25,16 (13,18)
H3K9(me2)	Asp135(OD2)	**Thr11(N)**	23,92	2.86 (0.08)	22.37 (12.47)
H3K9(me1)	Glu169(O)	**Lys9(N)**	76,68	2,86 (0.08)	19.95 10.11)
H3K9(me1)	Asp135(OD2)	**Thr11(OG1)**	64,96	2,69 (0,11)	15,59 (8,09)
H3K9(me1)	**Ser10(O)**	Lys241(NZ)	58,03	2,81 (0,09)	21,06 (10,38)
H3K9(me1)	**Gly13(O)**	Asn86(ND2)	44,24	2,88 (0,07)	19,07 (9,29)
H3K9(me1)	**Ala7(O)**	Glu169(N)	35,04	2,89 (0,07)	30,88 (11,70)
H3K9(me1)	Asp135(OD1)	**Thr11(OG1)**	33,71	2,71 (0,13)	23,20 (13,98)
H3K9(me1)	Asp135(OD2)	**Thr11(N)**	24,92	2,85 (0,08)	25,20 (14,98)

Hydrogen bonds analysis suggests that the interactions between JMJD2A and its substrates mainly involve main chain - side chain interactions. This finding contradicts with a previous work which states the majority of the interactions are main chain – main chain interactions [1]. Only Thr11 forms side chain-side chain interactions with Asp135 of JMJD2A. Lys9 forms main chain – main chain hydrogen bond with Glu169 of the enzyme and main chain – side chain hydrogen bond with Arg8. Backbone atoms of other residues except Gly12 of the substrates participate in hydrogen bonding with JMJD2A residues. Asn80, Asp 135, Glu169 and Lys235 of JMJD2A are common in hydrogen bond formation for three cases. The hydrogen bond between Asp311 of enzyme and Arg8 of substrate was not observed for H3K9(me1). Reconsidering the hydrogen bonding analyses we can conclude that there is no major difference between cases.

4.2 Binding Free Energy Calculations

It was reported by Whetstine at al[11] that JMJD2A was Lys (Me_3) specific enzyme and only had activity on Lys (Me_2) in the presence of excessive amount of enzyme and had no activity on monomethyllysine. In accordance with their findings Couture at al showed 20-fold reduced activity of this enzyme on dimethyllysine compared to trimethyllysine[3]. Based on these findings we thought that the activity specificity of JMJD2A may be related to binding affinity of the enzyme to its substrates. To determine whether there is a relation between activity specificity and binding affinity we computed the binding free energy for the three cases.

We computed the free energies according to the method named MM-PBSA that we explained in details in Methods section. In order to calculate the binding free energy we extracted snapshots from 4ns -16ns part of each trajectory due to the stability of H9K9(me3) tail in this part (See Figure 4.1 right panel). We determined the time interval of two tandem structures according to time correlation function of total energy. For each case we found out the time correlation function about 3. But bases on a previous study[38] we extracted the structures with a time interval of 12 ps resulting in 1000 snapshots. To be consistent with the force field, we used a value of 4 for internal dielectric constant. We did the calculations with single trajectory methods[28] and provided the results in Table 4.3.

Table 4.3 Binding free energy calculations results based on MM-PBSA. All values are in kcal/mol.

Systems	MM-PBSA	-T*S	$\Delta G_{binding}$
H3K9(Me_3)	-86,98	-65,94	**21,04**
H3K9(Me_2)	-82,05	-64,28	**17,77**
H3K9(Me_1)	-74,52	59,11	**15,41**

Activity of enzymes needs appropriate positioning of reactants which provided by formation of accurate interactions between enzymes and substrates. The state specificity of this enzyme may stems from binding affinity to its substrates. Our binding free energy results are consistent with the catalytic activity of JMJD2A on its substrates. According to this results JMJD2A has maximum affinity to H3K9(me3) and minimum affinity to H3K9(me1).

To understand the energetic determinants of substrate recognition at molecular level we need to consider the enthalpic and the entropic components of binding. Enthalpic contributions provide a measure of the strength of interactions between the receptor and the ligand. Entropic contributions involve the change in solvent entropy and the loss of solute conformational degrees of freedom.

The entropy term comprises translational, rotational, vibrational and configurational entropy of solute. In this study we computed translational, rotational and vibrational entropies by utilizing NMODE module of Amber8 and ignored the configurational entropy.

Due to the high computational demand, entropy calculations were performed only for a few snapshots which may cause a sampling problem. In order to overcome this problem we followed a new approach. In this approach firstly we clustered the snapshots in classes according to their similarities and provided a representative structure for each class[31]. Then we minimized each representative structure in gas phase using the conjugate gradient method until the root-mean-square of the elements of the gradient vector is less than 10^{-3} kcal $mol^{-1}A^{-1}$. For each minimized snapshot the entropy contributions of each component is computed at 300 K. After all, the entropy contribution of each cluster to the final entropy was obtained via weighting the relevant entropy value with the occurrence of the cluster.

To measure the reliability of MM-PBSA energy estimates we examined the convergence of enthalpic component of binding over the time and the results are provided in Figure 4.10.

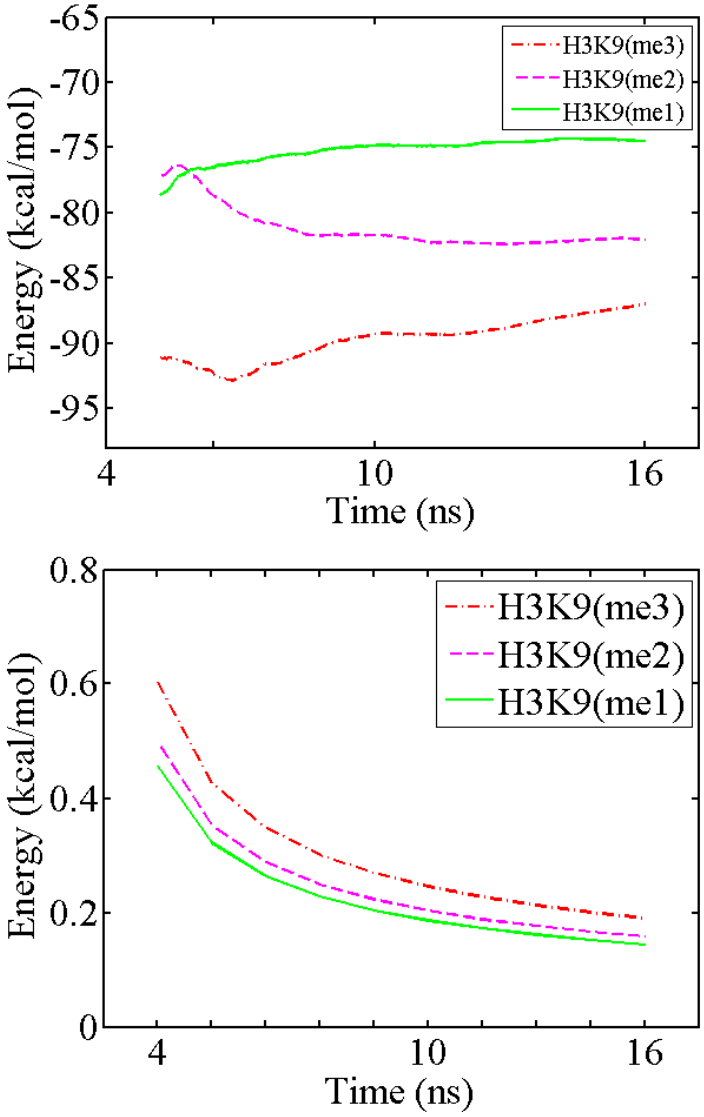

Figure 4.10 Convergence of the enthalpic component of binding, calculated with MM-PBSA at dielectric constant 4, assessed by the mean and the standard error for H3K9(Me_1), H3K9(Me_2) and H3K9(Me_3). H3K9(me1), H3K9(me2) and H3K9(me3) are shown in green, magenta and red colors respectively.

4.3 Binding Free Energy Decomposition

Binding free energy decomposition is a useful method which enables determination of contribution of each residue of both binding partners. This method provides insight into origin of binding. Since we aim to reveal state specificity of JMJD2A this method may help to determine the interaction types which anchor the peptides in substrate binding site. We perform binding free energy decomposition calculations for each case with 1000 structures that cover 4ns - 16 ns part of trajectory as well as MM-PBSA calculations. We tabulate results in Table 4.4 for enzyme residues and Table 4.5 for substrate residues.

Table 4.4: JMJD2A residues which are important in binding. Only the residues that make favorable contribution more than 1 kcal/mol are shown. Energy values are in kcal/mol.

Enzyme Residues	H3K9(Me1)	H3K9(Me2)	H3K9(Me3)
Tyr85	-1,76		
Asn86	-2,91	-1,34	-1,86
Asp135	-2,14	-2,29	-2,31
Ile168	-1,64	-1,62	
Glu169	-2,31	-2,66	-1,93
Gly170	-1,18	-1,21	-1,31
Val171	-1,41	-1,49	-1,15
Tyr175	-2,20	-2,82	-2,31
Tyr177	-1,05	-1,14	-1,73
Glu190			-1,43
His240	-1,41		
Lys241	-2,42	-1,50	-2,82
Met242	-2,07		-1,07
Thr289			-1,04
Asn290		-1,04	-1,02
Arg309		-3,43	-2,80
Asp311		-2,32	-3,10
Met312		-3,56	-1,07
Val313	-1,53	-1,29	-1,76

Table 4.5: Substrate residues which are important in binding. Only the residues that make favorable contribution more than 1 kcal/mol are shown. Energy values are in kcal/mol.

Substrate Residues	H3K9(Me1)	H3K9(Me2)	H3K9(Me3)
Ala7	-3,40	-3,67	-2,89
Arg8	-4,82	-8,21	-5,97
Lys9(me1/2/3)	-7,10	-12,7	-13,56
Ser10	-2,73	-2,86	-3,18
Thr11	-3,65	-5,07	-3,92
Gly12	-2,39	-2,17	-3,15
Gly13	-3,34	-2,25	-2,86
Lys14	-2,59	-3,43	-4,58

Binding free energy decomposition analyses revealed that Asn86, Asp135, Glu169, Gly170, Val171, Tyr175, Tyr 177, Lys241 and Val313 of JMJD2A make important contributions to binding free energy for three cases (see Table 4.4). Additionally for the case of H3K9(me3) Glu190, Met242, Thr289, Asp290, Arg309, Asp311 and Met312 of JMJD2A make important contribution to binding free energy. Ile168 and Val313 were reported to make van der Waals interactions with N-terminal residues of H3K9(me3) peptide while Asn86, His 240, Lys 241 and Met242 were found in same type of interactions with C-terminal of the substrate peptide[3]. Concurrently it was found out that Gly170, again Tyr177, Ser288 and Glu190 formed methylamonium binding pocket and Tyr175 Tyr177, Asp191 and Asn290 interacted with side chain of tri-meyhylated lysine. Our findings substantially correlate with experimental data except Asp191 and Ser288 [1-3].

All residues of substrate peptides seem important for binding. As expected the most contribution to binding free energy comes from modified Lys 9 although the values vary between cases. According to the binding free energy decomposition results Arg8 is

secondly important in binding. For both residues especially for modified Lys9 the large part of favorable energy contribution comes from van der Waals interactions which cause the difference between states. Electrostatic interactions are secondly important for these residues. In general this is the case for the residues which become prominent in binding.

4.4 Importance of Intra-substrate H-bonding

It was revealed that JMJD2A shows maximal activity with the H3K9(me_3) peptide substrate[2-3], suggesting that this sequence adopts an optimal conformation. The H3K9(me_3) substrate gains a broad 'W'-shaped conformation during binding. This bent peptide conformation is stabilized by intra-substrate hydrogen bond and required for sequence specificity of JMJD2A[2]. To determine whether intra-substrate hydrogen bonds occur we analyzed all potential hydrogen bond acceptors and hydrogen bond donors throughout the simulations. Here, to select hydrogen bonds we used the criteria hydrogen-bond distance, $R< 3.00$ A and hydrogen-bond angle, $120 < \theta < 180$. Contrary to expectations, we did not find out a hydrogen bond between H3 Ser10 side chain and main-chain H3 Gly12 which was reported by Ng at al. In their study they observed a strong reduction in activity of Ser10Ala mutant of H3K9me3 and related this finding to intra-substrate hydrogen bond. Unexpectedly, we discovered another intra-substrate hydrogen bond which was formed between H3 Arg 8 side chain and trimethyllysine main-chain (Figure 4.11). As shown in Figure 4.11, the side chain of Arg8 is twisted to establish interaction with main chain of trimethylysine. This hydrogen bond is occurred in a critical place where the first arc (Shown with A in Figure 4.11) is formed. Arc A seems important to direct the trimethyllysine into the methylammonium binding pocket. The intra-substare hydrogen bond creates a tensile force which may stabilize the necessary bent conformation of substrate peptide which directs the methylysine side chain into the binding pocket. Throughout the simulation the W-shaped of substrate was conserved which is indicative of stability. This stable conformation is most probably outcome of the intra-substrate

hydrogen bond as well the interactions between enzyme and substrate. Based on this finding we proposed that Arg8 has a critical role in stabilizing the substrate. To further assess this claim we perform computational alanine scanning mutagenesis. We mutated Agr8 to Ala and performed MD simulation of 12 ns. With the resulting trajectory we performed some analyses. Compared to the wild type, we expected the mutant substrate to be more mobile. Based on this, we measured FE-NZ distance for both systems and provided the results in Figure 4.10. We defined the mobility as the number of picks higher than 6.0 Å. In this respect we found out the mobility 4.8/ns and 12.1/ns for wild type and mutant respectively. In accordance with our expectations we observed a difference between those two systems. Thereafter we calculated average structure of bonded peptide for both the wild type and the mutant. From comparison of these two structures we found out that the mutant has much broader 'W' shape especially at the point of arc A which implies the significance of Arg residue at this position.

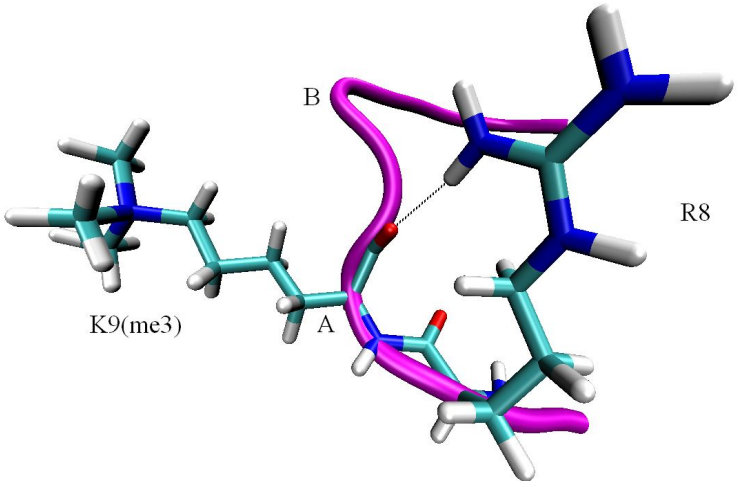

Figure 4.11: Representative Snapshot of H3 Arg8 side chain and H3 Lys (Me_3) main chain hydrogen bond. The backbone of H3 tail is shown in magenta and the two arcs of W-shaped conformation are indicated with A and B. Hydrogen bonds are defined by the distance less than 3Å and donor-hydrogen-acceptor angle grater than 120^o.

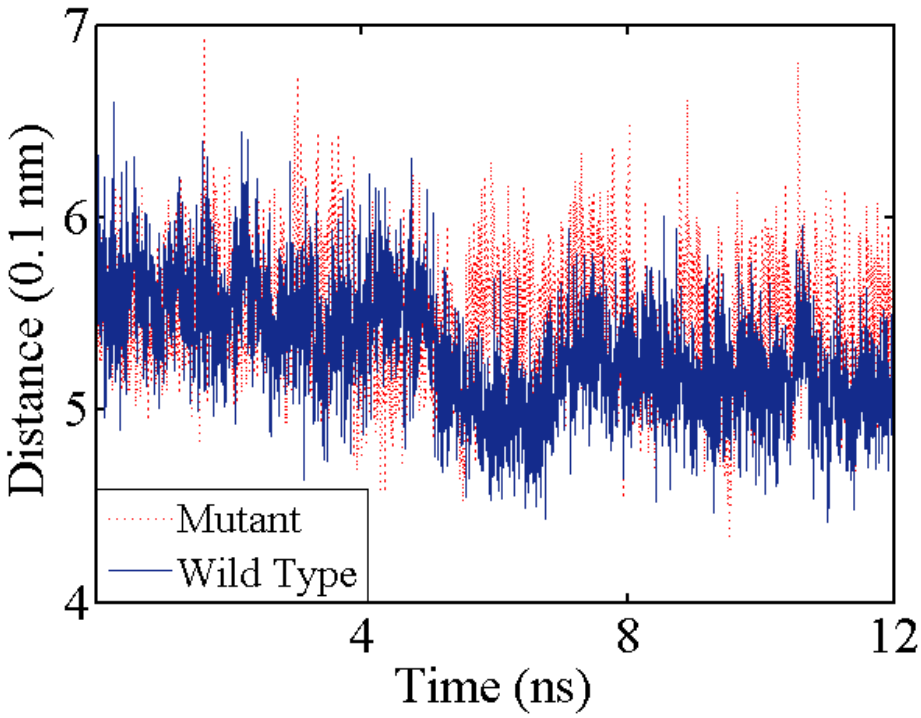

Figure 4.12: The distance between NZ and Fe(II) ion was measured for both wild type and mutant. The number of picks that higher than 6 Å per second was calculated as a measurement of mobility.

To assess the issue from the energy point of view we performed computational alanine scanning mutagenesis[33]. In this approach, basically we mutated Agr8 to Ala via truncation of side chain atoms and compute the binding free energy difference between the wild type and mutant. For H3K9(me_3) case for the enthalpic component of binding reduces with an amount 6,43 kcal/mol (see Table 4.6 Method I) . We also performed 16 ns production of MD for the mutant and computed the binding free energy difference. The results are provided in Table 4.6 Method II. The entropy component of binding was not

calculated since it was assumed, based on a previous work, that its contribution to $\Delta\Delta G_{binding}$ is negligible[39]. Both structural and energetic analyses support the importance of Arg 8 in binding.

Table 4.6: Computational alanine scanning mutagenesis results

	MM/PBSA			MM/GBSA		
	Wilde type	**Mutant**	$\Delta\Delta G_{binding}$	**Wilde type**	**Mutant**	$\Delta\Delta G_{binding}$
Method I	86,98	-80,55	**6,43**	90, 96	-80,49	**10,47**
Method II	86,98	-84,63	**2,35**	90, 96	-83,48	**7,48**

Chapter 5

CONCLUSION

In this work we have presented the results of our all-atom MD simulations of JMJD2A in complex with its substrates; H3K9(me1), H3K9(me2) and H3K9(me3). Our object of interest was the methylation state specificity of JMJD2A. In total our simulation time reached 20 ns with 2 ns of equilibration phase for each case. We performed some structural, hydrogen bonds, binding free energy, alanine scanning and binding free energy decomposition analyses and provide those results in previous section. In this section we are going to discus those results and query their validity.

JMJD2A has two metal centers which we parameterized according to the method explained in previous sections. The method worked for Fe(II) center very well while for Zn(II) we needed to put harmonic constrain to keep the metal around its original position. The distances between metal and its coordinating atoms seem in good agreement with crystallographic structures. The lower values of mean fluctuations of Fe(II) over 18 ns with respect to its initial position are indicative of successive parameterization. It was interesting to look at the behavior of water molecule that coordinated iron kation for each case. For three cases in any time one water molecule was found in bond distance to Fe(II). We examined the behavior of water molecules which were found in certain positions in crystallographic structure. For H3K9(me1) one water molecule (wat3) kept its initial position for 10 ns the other molecule (wat2) remains around its initial positions over 18 ns

as well as the case of H3K9(me2). Wat3 kept its position via forming hydrogen bonds with Ser288 and Gly177 main chain oxygens whereas wat1 formed hydrogen bonds with Ser288 and Tyr171 side chains. These finding are in consistent with crystallographic studies.

We further assessed the presence of water molecules in certain positions via examining the behavior of cationic head of modified lysine and the positions of methyl groups with respect to Fe(II) for each case. The methyl groups did not locate randomly and had three options to sit. Compared to H3K9(me3) case, H3K9(me2) and H3K9(me1) cationic heads were much more restricted which was consistent with the presence of water molecules. We demonstrated that restriction on the angular motion in dimethyllysine and monomethyllysine cases is an outcome of avoidance of steric overlaps. Since methylammonium has symmetry around the NZ atom of methyllysine, 120^{o} rotation of methylammonium head results in the same orientation. The equivalency of the occupied volume by each methyl group prevents the atoms from any steric overlaps during rotation.

The occurrence of methyl groups in a certain proximity of Fe(II) was in good agreement with presence of water molecules and the orientation of methyl groups. For H3K9(me3) both the presence of three methyl groups and the absence of water molecules which may restrict orientation of methyl groups raise the occurrence of methyl groups in close proximity of Fe(II). Also the increase in atom numbers in cationic head might raise the van der Waals interaction strength between the head and iron ion.

Hydrogen bonding analysis suggests that the interaction between JMJD2A and its substrates mainly involves main chain - side chain interactions. This finding contradicts with a previous work which states the majority of the interactions are main chain – main chain interactions[1]. Reconsidering the hydrogen bonding analyses we can conclude that there is no major difference between cases.

Activity of enzymes needs appropriate positioning of reactants which provided by formation of accurate interactions between enzymes and substrates. The state specificity of

this enzyme may stems from binding affinity to its substrates. To asses this issues we performed binding free energy analysis with MM-PBSA method and to reveal out the critical residues we carried out binding free energy decomposition which based on MM-GBSA. As demonstrated in previous section our binding free energy results from MM-PBSA are consistent with the catalytic activity of JMJD2A on its substrates.

Binding free energy decomposition analyses revealed that Asn86, Asp135, Glu169, Gly170, Val171, Tyr175, Tyr 177, Lys241 and Val313 of JMJD2A make important contributions to binding free energy for three cases. Additionally for the case of H3K9(me3) Glu190, Met242, Thr289, Asp290, Arg309, Asp311 and Met312 of JMJD2A make important contribution to binding free energy. Ile168 and Val313 were reported to make van der Waals interactions with N-terminal residues of H3K9(me3) peptide while Asn86, His 240, Lys 241 and Met242 were found in same type of interactions with C-terminal of the substrate peptide(1). Concurrently it was found out that Gly170, again Tyr177, Ser288 and Glu190 formed methylamonium binding pocket and Tyr175 Tyr177, Asp191 and Asn290 interacted with side chain of tri-meyhylated lysine. Our findings are in good agreement with experimental data[1-2].

All residues of substrate peptides seem important for binding. As expected the most contribution to binding free energy comes from modified Lys9 although the values vary between cases. According to the binding free energy decomposition results Arg8 is secondly important in binding. For both residues especially for modified Lys9 the large part of favorable energy contribution comes from van der Waals interactions which cause the difference between states. Electrostatic interactions are secondly important for these residues. In fact this can be generalized for all residues on enzyme- substrate interface.

Arg8 is important in two ways. First it is secondly important in binding and it forms intra-substrate hydrogen bond with modified lysine in H3K9(me3) and H3K9(me2). As we explained in previous section this intra-substrate hydrogen bond may stabilize the

necessary conformation of substrate peptide. To further asses the role of Arg8 in binding we performed computational alanine scanning analysis. The mutation reduced the binding free energy with a amount of 2,35 to 10,40 kcal/mol according to the method used (See Table 4.6). Additionally, structural analyses support the significance of Arg8 in this position.

In general we performed MD simulations of three systems; JMJD2A in complex with H3K9(me1), H3K9(me2) and H3K9(me3) to reveal the state specificity of JMJD2A. With this object of interest we performed some structural and energetic analyses and provided results.

BIBLIOGRAPHY

1. Chen, Z., et al., *Structural basis of the recognition of a methylated histone tail by JMJD2A*. Proc Natl Acad Sci U S A, 2007. **104**(26): p. 10818-23.
2. Ng, S.S., et al., *Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity.* Nature, 2007. **448**(7149): p. 87-91.
3. Couture, J.F., et al., *Specificity and mechanism of JMJD2A, a trimethyllysine-specific histone demethylase.* Nat Struct Mol Biol, 2007. **14**(8): p. 689-95.
4. Derewenda, Z.S., Lee, L. & Derewenda, U., *The occurrence of C-H···O hydrogen bonds in proteins*. J. Mol. Biol., 1995. **252**: p. 248–262.
5. Bird, A., *Perceptions of epigenetic.* Nature, 2007. **447**: p. 396-398.
6. Kundu, T.K.D., Dipak, *Chromatine and Disease*. Subcellular Biochemistry, ed. J.R. Harris. Vol. 41. 2007, Wooster, Ohio: Springer.
7. Santos-Rosa, H. and C. Caldas, *Chromatin modifier enzymes, the histone code and cancer.* Eur J Cancer, 2005. **41**(16): p. 2381-402.
8. Mattei M.G., Luciani J. *Heterochromatin, from Chromosome to Protein*. Atlas of Genetics and Cytogenetics in Oncology and Haematology 2003.
9. Moss, T.J. and L.L. Wallrath, *Connections between epigenetic gene silencing and human disease.* Mutat Res, 2007. **618**(1-2): p. 163-74.
10. Chen, Z., et al., *Structural Insights into Histone Demethylation by JMJD2 FamilyMembers.* Cell, 2006. **125**: p. 691–702.
11. Whetstine, J.R., et al., *Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases.* Cell, 2006. **125**(3): p. 467-81.

12. Karlin, S., Zhu, Z.Y. and Karlin, K.D. *The extended environment of mononuclear metal centers in protein structures*. Proc Natl Acad Sci U S A, 1997. **94**(26): p. 14225-30.
13. Jorgensen, W.L., *Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water*. J. Am. Chem. Soc., 1981. **103**: p. 335-340.
14. Duan, Y., et al., *A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations*. J Comput Chem, 2003. **24**(16): p. 1999-2012.
15. Case, D. A., Darden, T.A., Cheatham, T.E., Simmerling, C.L., Wang, J. and Duke, R.E., *AMBER 10*. 2008, University of California: San Francisco, CA.
16. *Discovery Studio Visualizer*, Accelrys Inc: San Diego, CA, USA.
17. Frisch, M.J.T., Schlegel, G.W., Scuseria H.B., Robb G.E., Cheeseman M.A. and Montgomery J.R., *Gaussian 03*. 2003, Gaussian Inc: Pittsburgh, PA.
18. Bayly C. I., Cieplak P., Cornell W. D., Kollman P. A., *well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model*. J. Phys. Chem, 1993. **97**: p. 10269–10280.
19. Pigache A., Cieplak P.C., Dupradeau F.Y. *Automatic and highly reproducible RESP and ESP charge derivation: Application to the development of programs RED and X RED*, in *227th ACS National Meeting*. 2004: Anaheim, CA, USA.
20. Wang J., W.R.M., Cardwell J.W., Kollman P.A., Case A. D., *Development and Testing of a General Amber Force Field*, in *Journal of Computational Chemistry*. 2004. p. 1157–1174.
21. Peraro M. D., S.K., Lamoureux G, De Vivo M., and K.M.L. DeGrado W. F., *Modeling the charge distribution at metal sites in proteins for molecular dynamics simulations*. Journal of Structural Biology 2006. **157**: p. 444–453.

22. Darden T, Y.D., Petersen L., *Particle Mesh-Ewald-an N Log(N) method for Ewald sums in Large systems* J Chem Phys, 1993. **98**: p. 10089-92.
23. Kale L., Skeel R., Bhandarkar M., Brunner R., Gursoy A., Krawetz N., Phillips J., Shinozaki A., Varadarajan K., Schulten K., *NAMD2: Greater scalability for parallel molecular dynamics*. J. Comput. Phys., 1999. **151**: p. 283-312.
24. Phillips, J.C., et al., *Scalable molecular dynamics with NAMD*. J Comput Chem, 2005. **26**(16): p. 1781-802.
25. Martyna G.J. Tobias, Klein M.L., *Constant pressure molecular dynamics algorithms*. J. Chem. Phys, 1994. **101(5)**: p. 4177-4189.
26. Feller S.E., Zhang Y., Pastor R.W. and Brooks B.R., *Constant pressure molecular dynamics simulation: The Langevin piston method.* J. Chem. Phys, 1995. **103(11)**: p. 4613-4621.
27. Miyamoto S., Kollman P.A., *SETTLE: An analytical version of the SHAKE and RATTLE algorithm for rigid water models.* 1992. **13**: p. 952–962.
28. Kollman, P.A., et al., *Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models.* Accounts of Chemical Research, 2000. **33**: p. 889-897.
29. Sitkoff D.S., Sharp K. A.; Honig B., *Accurate Calculation of Hydration Free Energies Using Macroscopic Continuum Models*. J. Phys. Chem., 1998. **98**: p. 1978-1983.
30. Rareya M., Kramera B., Lengauera T and Klebe G, *A Fast Flexible Docking Method using an Incremental Construction Algorithm.* J Mol Bio, 1996. **261**(3): p. 470-489.
31. Jianyin Shao, S.W.Tanner., Nephi Thompson and Thomas E. Cheatham, *Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms.* J. Chem. Theory Comput, 2007. **3**: p. 2312-2334.

32. Case, D.A.C., T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, and A.S. K. M.; Onufriev, C.; Wang, B.; Woods, R. J., *The Amber biomolecular simulation programs*. J. Comput. Chem, 2005. **26**: p. 1668-1688.

33. Kollman, I.M.a.P.A., *Computational Alanine Scanning To Probe Protein-Protein Interactions: A Novel Approach To Evaluate Binding Free Energies.* J of American Chemical Society, 1999. **121**: p. 8133-8143.

34. Holger Gohlke, C.K.a.D.A.C., *Insights into Protein–Protein Binding by Binding Free Energy Calculation and Free Energy Decomposition for the Ras–Raf and Ras–RalGDS Complexes*. J Mol Bio, 2003. **330**: p. 891–913.

35. Onufriev A., Bashford, D., Case, D. A. *Modification of the Generalized Born Model Suitable for Macromolecules*. Vol. 104. 2000.

36. Jana M. Simmons, T.A.M.a.R.P.H., *FeII/a-ketoglutarate hydroxylases involved in nucleobase, nucleoside, nucleotide, and chromatin metabolism.* Dalton Transactions, 2008: p. 5132–5142.

37. Couture, J.F., et al., *Catalytic roles for carbon-oxygen hydrogen bonding in SET domain lysine methyltransferases.* J Biol Chem, 2006. **281**(28): p. 19280-7.

38. Wan, S., P.V. Coveney, and D.R. Flower, *Peptide recognition by the T cell receptor: comparison of binding free energies from thermodynamic integration, Poisson-Boltzmann and linear interaction energy approximations.* Philos Transact A Math Phys Eng Sci, 2005. **363**(1833): p. 2037-53.

39. Huo, S., I. Massova, and P.A. Kollman, *Computational alanine scanning of the 1:1 human growth hormone-receptor complex.* J Comput Chem, 2002. **23**(1): p. 15-27.