Molecular Recognition of the Methylated Histone Peptides by the Tandem Tudor Domains of JMJD2A

by

Musa Özboyacı

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

September 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

Musa Özboyacı

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:

_____

Özlem Keskin, Ph. D. (Advisor)

_____

Attila Gürsoy, Ph. D.

_____

Burak Erman, Ph. D.

Date: _____

ABSTRACT

In this thesis, we report a detailed molecular dynamics simulation and MM-PBSA/GBSA approach (MM: Molecular Mechanics; PB: Poisson Boltzmann; GB: Generalized Born; SA: Surface Area) analysis, unraveling the recognition of the methylated histone tails H3K4me3, H4K20me3, H4K20me2 and H3K9me3 by JMJD2A-tudor. In this respect, 25 ns fully unrestrained molecular dynamics simulations were conducted for each of the bound and free structures. We investigated the important hydrogen bonds and coulombic interactions between the tudor domains and the peptide molecules; hence unveiled critical residues occupied in stabilizing the complexes. Normal mode and molecular mechanics calculations were performed to obtain the entropic determinants of the binding affinities. Suggested by the resulting binding free energies obtained via GB and PB approaches, we found that H4K20me3 peptide has the highest affinity to JMJD2A-tudor in GB calculations whereas H4K20me2 peptide has the highest affinity to JMJD2A-tudor in PB calculations. Furthermore, we discerned that H3K9me3 peptide has the lowest affinity to JMJD2A-tudor in both of the model calculations. We also revealed that while H4K20me2 peptide adopting the same binding mode with H4K20me3 peptide, H3K9me3 peptide adopts the same binding mode with H3K4me3 peptide. Decomposition of the enthalpic and the entropic contributions to the binding free energies indicated that the recognition of the histone peptides is mainly driven by favourable van der Waals interactions in both GB and PB models. Based on GB calculations pairwise and per residue decomposition of the binding free energies with backbone and sidechain contributions as well as their energetic constituents were also carried out to identify the hotspots of the structures. Thus, the van der Waals and the electrostatic interactions which are prominent for the recognition of the peptides were clarified.

ÖZET

Bu çalışmada, moleküler dinamik simülasyonu ve MM-PBYA/GBYA (MM: Moleküler Mekanik; PB: Poisson Boltzmann; GB: Genellenmiş Born; YA: Yüzey Alanı) yöntemiyle, metillenmiş histon kuyruklarının (H3K4me3, H4K20me3, H4K20me2 ve H3K9me3) JMJD2A tudor domeni tarafından tanınması incelenmiştir. Bu bağlamda, kompleks ve serbest yapıların 25 ns lik moleküler dinamik simülasyonları yapılmıştır. Kompleks yapıların stabilizasyonunu sağlayan kritik amino asitlerin tespit edilmesi amacıyla tudor domeni ve peptitler arasındaki hidrojen bağları ve tuz köprüleri incelenmiştir. Bağlanma afinitelerinin entropik etkilerini belirlemek amacıyla normal mod analizi ve moleküler mekanik hesaplamaları gerçekleştirilmiştir. GB ve PB uygulamalarından elde edilen bağlanma enerjileri sonuçlarına dayanarak, H4K20me3 peptidinin GB hesaplamalarında en yüksek afiniteye sahip olduğu, H4K20me2 peptidinin ise PB hesaplamalarında en yüksek afiniteye sahip olduğu tespit edilmiştir. Bununla birlikte, H3K9me3 peptidinin her iki model hesaplamalarında JMJD2A-tudor domeni için en düşük afiniteye sahip olduğu görülmüştür. Ayrıca, H4K20me2 peptidinin H4K20me3 peptidiyle aynı bağlanma şekline sahip olduğu ve H3K9me3 peptidinin de H3K4me3 peptidiyle aynı bağlanma şeklini benimsediği görülmüştür. Bağlanma enerjilerinin entalpik ve entropik analizi, histon peptitlerinin tanınmasının her iki model için de (GB ve PB) genellikle van der Waals etkileşimleri tarafından gerçekleştirildiğini göstermiştir. Bununla birlikte GB hesaplarına dayanarak, yapılardaki kritik amino asitlerin belirlenmesi amacıyla, amino asitlerin ikili ve tekli ana zincir ve yan zincir katkılarının yanısıra enerjetik bileşenleri analiz edilmiştir. Böylelikle peptidlerin tanınması için önemli olan elektrostatik ve van der Waals etkileşimleri açıklığa kavuşturulmuştur.

ACKNOWLEDGEMENTS

First I would like to thank my advisors, Dr. Özlem Keskin, Dr. Attila Gürsoy and Dr. Burak Erman for helping, guiding and supporting me in my studies and teaching me how to study independently. I deeply appreciate them for orienting me to the exciting area of molecular dynamics simulation studies.

I would like to thank my research partners Özlem Ulucan and Deniz Şanlı for helping, supporting and motivating me. Thank you for sharing your time and your friendship, which mean a lot to me.

I also would like to thank my friends Müge Karaman, Abdullah Murat Turan and Ömer Kırkağaçlıoğlu whom I started to the Computational Sciences and Engineering program together with. I will never forget the days we tried to catch up on our assignments just on time.

I would like to thank my colleagues Gözde Eskici and Mehmet Ali Öztürk for their invaluable friendships.

Having spent two wonderful years together, I would like to express my appreciation to Seda Yaşar (together with her family), Özge Narin, Esra Satıcı and Serra Caner. It was really fun to enjoy the parties and the meetings that we participated together. More importantly, I appreciate your invaluable friendships.

I also thank Turan Bulmuş and Hakan Doğan for their neighborhood as well as for their friendships. Thank you guys a lot for supplying food to me and Çağrı in case of emergency and for the fun we had while playing cards together.

For being such a great housemate and friend, I deeply thank Mehmet Çağrı Dedeoğlu. I am really grateful for enduring me for two years with considerable kindness.

My special thanks go to Nihan Çömden for being such a good, kind and considerate friend. I will never forget the times we have spent together.

Finally and most importantly, I would like to express my gratitude to my family members for always being there for me. I really thank my sister Emine and my brother Mehmet for always supporting me and providing love to me. I am extremely grateful to my Mother and my Father for the affection, support and love throughout my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

JMJD2A is a histone lysine demethylase enzyme which plays a prominent role in the development of prostate and esophageal squamous cancers [1]. Consisting of a JmjC, a JmjN, two PHD and two tandem tudor domains, JMJD2A recognizes and binds to four different methylated histone peptides: H3K4me3, H4K20me3, H4K20me2 and H3K9me3, via its tudor domains [2]. Of the four histone peptides, only recognition of the H3K4me3 and H4K20me3 by JMJD2A-tudor has been identified. Moreover, it has been revealed that two distinct binding modes were adopted by H3K4me3 and H4K20me3 peptides upon complex formation [3]. However, complete picture in understanding the recognition of the four of the methylated peptides is still unclear.

To clarify the missing points in previous studies and to gain insights into the recognition of other structures of bound JMJD2A-tudor (to H4K20me2 and H3K9me3), whose structures have not been revealed yet, in this study, we investigated the dynamics of JMJD2A-tudor to understand the mechanism of recognition of the specific methylated peptide patterns. To accomplish this, 25 ns molecular dynamics simulations of the four structures of JMJD2A-tudor liganded to H3K4me3, H4K20me3, H4K20me2 and H3K9me3, as well as the free tudor domains and the free peptide ligands were performed.

Employed in the molecular dynamics simulations, missing parameters for the trimethyllysine and dimethyllysine residues were generated using quantum mechanical techniques. For JMJD2A-tudor-H4K20me2 complex, initial structure was taken same as

the trimethylated structure in which the trimethyl group was turned into a dimethyl group. Since there was not any available structural information on binding mode of JMJD2A-tudor-H3K9me3 complex, initial structure of the complex was modelled using docking simulations. After the simulations, binding free energies using molecular mechanics Poisson-Boltzmann/generalized Born surface area (MM-PBSA/GBSA) approaches were performed and based on the outcomes, critical residues were identified.

As a result of this study, we unravelled and explained several characteristics of the recognition of the peptides by the tudor domains of JMJD2A.We showed that the binding mode of H3K9me3 is same with H3K4me3 and the residues that are involved in recognition of these two peptides are very similar. We clarified the remarkable interactions between the receptor and the peptide. Furthermore, we identified new residues such as Ser938 and Glu929 involved in strong interactions with the peptides. Being energetically favourable upon binding, we identified the residues in the complex structures which play fundamental role in complexation. Along with residues Asp945, Asn940 and Asp939 of the protein, we represented new hot spots such as Ser936, Phe937 and Asp969.

Chapter 2 provides the background information about the tudor domains of JMJD2A along with an overview of histone code hypothesis and JMJD2A. Significance and basics of the computational tools used in this study are summarized.

Chapter 3 describes the computational methods used in this study in detail. Procedures and details of initial system preparations, homology modelling of the missing residues and parameterization of the non standard residues are given. Detailed information on theories of molecular docking and molecular dynamics simulations is extended along with their procedures in this study. Calculations of enthalpic and entropic components of the binding free energy are given with a brief theoretical background.

In chapter 4, convergence and stability analysis of the simulation trajectories and energy calculations are introduced with statistical results.

Chapter 5 gives the structural details of the systems investigated in this study. Critical interactions and comparisons between the structures are discussed in detail.

In chapter 6, MM-PBSA/GBSA calculation results are provided for each system with energetic contributions revealed by the decomposition of the energy values. In addition, identified as favouring the binding energetically, hot spots are discussed in this chapter. Detailed comparative discussions are extended for each part.

The thesis is concluded with a concise summary of the work done in this study.

Chapter 2

LITERATURE REVIEW

2.1 Overview

Chromatin fibre in eukaryotes, which is composed of DNA and proteins, is responsible for the packaging of the DNA to fit its size in the nucleus. Chromatin is composed of repeating nucleosome core particles. Nucleosome particles in the chromatin basically consist of an octamer structure of core histone proteins and the DNA molecule wrapped around it. The DNA molecule around the core is made up of nearly 146 pairs of nucleotides and each octameric core is composed of H2A, H2B H3 and H4 proteins (Figure 2.1) [4][5][6][7]. Nucleosomes also contain a linker histone protein H1, which is responsible for the stability of the nucleosome [5]. Because histone proteins are highly conserved in eukaryotes, they were thought to be the simple repeating units of chromatin fibre [8]. However, in the early nineties, it was discovered that histone proteins were playing a crucial role in many biological events, such as signalling, directing the proteins into DNA, transcription activation or deactivation and so on [9][10]. This biological functioning is achieved by means of covalent modifications (acetylation, methylation, phosphorylation, sumolation, ubiquitination, vs.) of the histone molecules [11]. Today, we know that the combination of the covalent modifications in the histone tails form a histone code, which leads to different biological processes. Among all the covalent modifications that take place in histone tail, methylation and so demethylation have a considerable role in transcriptional

regulation and chromatin condensation [12]. It has been proposed that methylation of lysine residues in H3 and H4 histone proteins, specifically give rise to key downstream chromatin-related functionalities [5][10][13][14]. Researchers once considered that histone lysine methylation was irreversible and could be shifted with a new histone molecule that has not been methylated thus erasing the methyl mark on the protein [1][15][16]. However, recent studies showed that there were some other mechanisms for the removal of the methyl marks from the lysine residues of the histone.

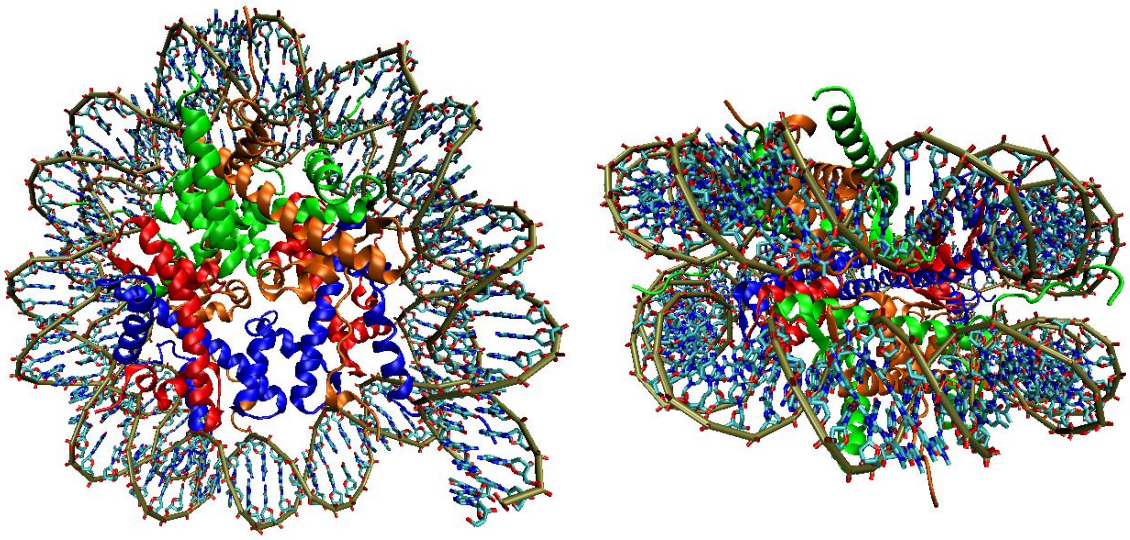


Figure 2.1: Crystal structure of a nucleosome composed of H2A (orange), H2B (green), H3 (Blue) and H4 (Red) proteins and DNA molecule wrapped around the histone proteins.

The discovery of histone lysine demethylases (HDMs) has demolished the previous idea of irreversibility of histone lysine demethylation. So far, only two families of histone lysine demethylase enzymes have been discovered. First family, Lysine Specific

Demethylase (LSD1) is a flavin dependant enzyme, which specifically demethylases mono and di methylated Lys4 and Lys9 of H3 tail [16][17]. After identification of LSD1, a second family of HDMs; JmjC has been discovered. The JmjC HDMs are α-ketoglutarate dependant oxygenases, which catalyse the demethylation reaction of di and tri methylated Lys9 and Lys36 of H3 tail [16][17]. Since the JmjC family of HDMs has been discovered lately and is relatively new, there are still many questions about the functioning of the enzymes that have not been answered yet. Therefore we put emphasis on the importance of finding new breakthroughs of this enzyme family.

The JMJD2 is the first family that has been identified among the several other members of the family of JmjC. Recent protein interaction studies showed that JMJD2 family enzymes are occupied in cancer related functionalities. Therefore, this family of enzymes, show promises as cancer targets [1]. Since many structural studies of JMJD2 family of enzymes have been performed, we decided to concentrate on the JMJD2A enzyme, which is a member of the JMJD2 family.

The JMJD2A enzyme has shown to have a pivotal role in esophageal squamous and prostate cancers [1]. It has been identified as one part of the complex N-CoR that has repression activity [9]. As a component of the corepressor complex, JMJD2A binds to the Androgen receptor (AR) and Retinoblastoma protein (Rb) thus initializing downstream effects [18][19].

The JMJD2A is basically composed of four different domains: JmjC, JmjN, 2 PHD and 2 tandem tudor domains (Figure 2.2). The catalytic site of the enzyme includes JmjC and JmjN domains (Figure 2.3). The demethylation reaction, where succinate and formaldehyde are generated to remove methyl groups, takes place in the catalytic site of the enzyme with the presence of Fe(II) and α-ketoglutarate [1][17][20].

Figure 2.2: Domains of JMJD2A, a JmjN a JmjC , two PHD and two tandem tudor.

PHD and Tudor domains of the JMJD2A enzyme are supposed to be occupied in the recognition. The recognition of specific patterns on the chromatin directs the enzyme to methylated lysine residues on the histone tail, thus leading to specific changes in the histone code [15]. So far, there has not been any experimental study on the specificity and the recognition mechanism of JMJD2A PHD domains. Therefore, the question of how PHD domains function in the enzyme still remains to be answered.

Tudor domains are a member of Tudor royal family, with other similar functioning domains including Chromo, plant Agenet, MBT and PWWP domains. Like chromodomains, Tudor domains are mainly responsible for the interaction with the methylated lysine and arginine residues in the proteins [5][21]. Interaction mainly occurs within the side chains of the Tudor domains and the side chains of the methylated residues. Tudor domains of JMJD2A bind mostly to trimethylated H3K4, trimethylated H3K9 and di,trimethylated H4K20. These interactions require Suv39h enzymes, which are responsible for transcriptional repression, thus supporting the finding that JMJD2A protein is a component of N-CoR corepressor complex [2]. The Tudor domains interact with different histone tails by different binding modes. It has been previously demonstrated that point mutations on these domains alter the functionality by repressing the recognition of one pattern but do not change the other [3]. Understanding the underlying specificity of the difference in the recognition of the domains is important since selective targeting for the Tudor domains may be achieved via designing selective drugs.

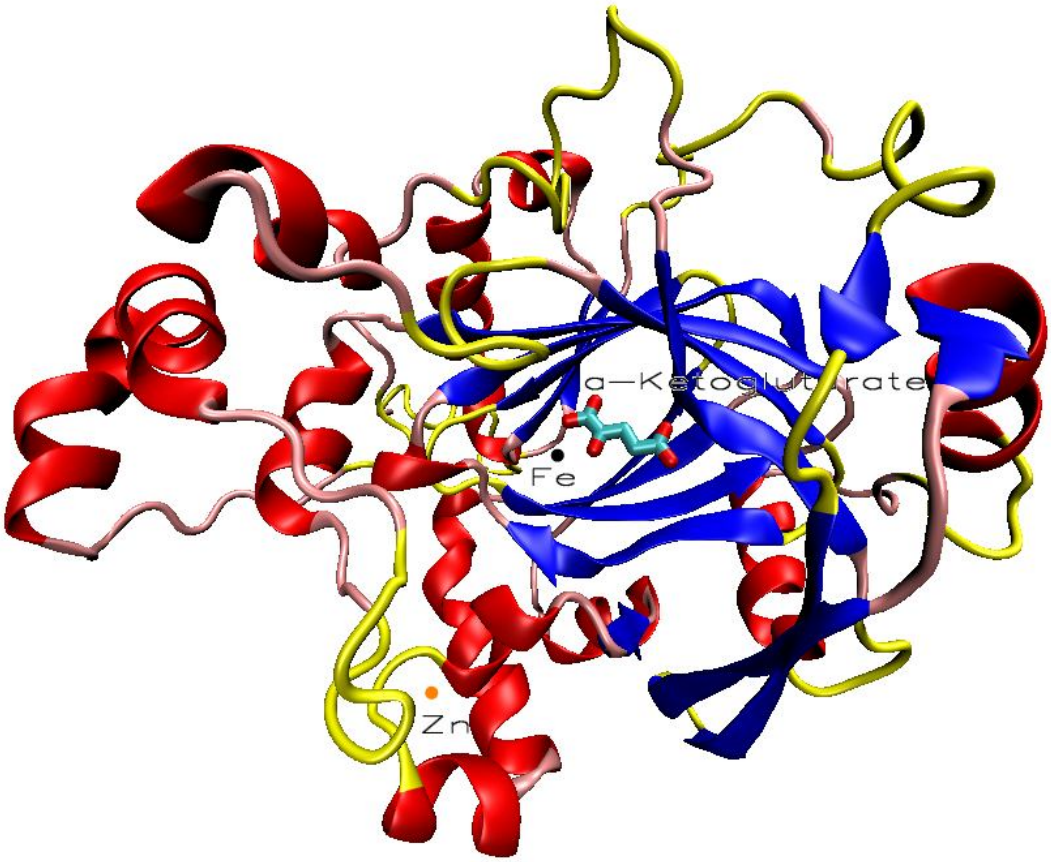


Figure 2.3: Catalytic site of JMJD2A consisting of JmjN and JmjC domains. Catalytic site includes a Zinc finger motif with a Zn(II) ion and a Fe centre composed of a Fe(II) ion and an α-ketoglutarate molecule.

2.2 JMJD2A-Tudor Domain

The double Tudor domains in the JMJD2A are tandem and bilobal. The tandem Tudor domains have a saddle shaped structure in which each lobe interweave with each other (Figure 2.4). Lobes in the tandem Tudor domains are named as hybrid Tudor domain 1 and 2 (HTD 1-2). Methylated peptide only binds to a specific crevice of HTD2 [3]. Compared

to HTD1, HTD2 is more negatively charged on the surface, so that it facilitates the binding of the methylated residues that are positively charged [21]. The structure of the tandem Tudor domains of JMJD2A differs than other proteins such as double Tudor domain of 53BP1, where each domain is folded independently [22]. The structural difference might suggest that each of the lobes in the tandem Tudor domains of JMJD2A strongly affect the dynamics and functioning of the other.

JMJD2A Tudor domains interact with trimethylated H3K4, trimethylated H3K9 and di,trimethylated H4K20. In mammals, methylation of H3K4 is mostly associated with transcriptional activation, thus antagonizing the effect of the methylation of H3K9 and H3K36 whereas methylation of H4K20 is associated with gene silencing [23]. As a consequence; demethylation reaction gives birth to the opposite regulations. Since JMJD2A functions mostly in multimeric form, different combinations of interactions with methylated H3K4, H3K9 and H4K20 might be targeting the enzyme to different destinations [3].

Previous studies have shown how two methylated peptides (H3K4me3 and H4K20me3) interact with JMJD2A-Tudor, although they do not share a similar amino acid sequence but methylated lysine residues. Interactions of methylated peptides with JMJD2A-Tudor occur in different binding modes by nearly same K_d values [3][21]. Although many of the residues that are occupied in the interactions are identified, how structural and dynamical behaviour of the protein consequences the modal difference in binding is still not well understood.

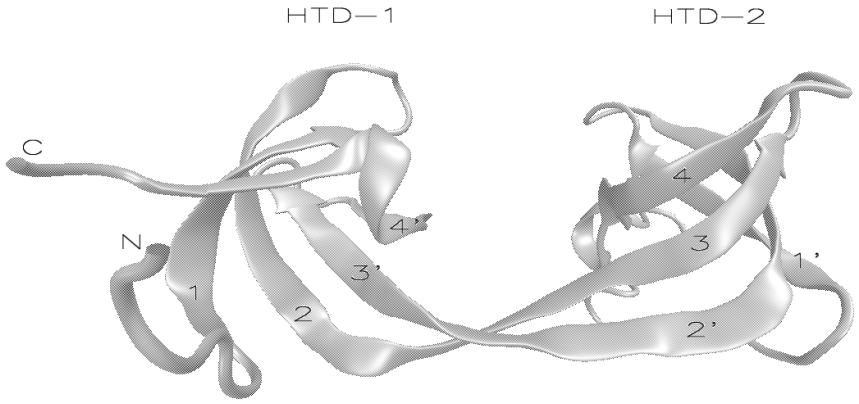


Figure 2.4: Secondary structure of JMJD2A-tudor domain. The tandem hybrid tudor domains have an interdigitated structure in which structural motifs are exchanged between each other. β2 and β3 strands are swapped between the hybrid domains. The two lobes of the structure are named as Hybrid Tudor Domain 1 (HTD-1) and Hybrid Tudor Domain 2 (HTD-2).

2.3 Homology Modelling

Recent improvements in computer science and numerous experiments that have been performed resulted in the development of new incomes in determination of tertiary structures of proteins [24]. There is a lot of interest concentrated on the prediction of the protein structure in 3D, especially in the studies of site-directed mutagenesis, structural studies and docking studies of proteins [25]. However, amino acid sequences obtained from the genome projects are much more than 3D structures that have been obtained via NMR or X-ray crystallography. The reason why this is so is that NMR and NMR and X-ray techniques are quite time consuming with the chance of failure, thus leading to the requirement of some novel techniques [26][27].

Among the other methods, Homology modelling is an effective tool to determine the tertiary structure of a protein [28]. Homology modelling essentially makes use of the three dimensional structures of the proteins having similar amino acid sequence to determine the structure of other proteins. Similarity of amino acid sequences are elevated in homologues proteins. Since the tertiary structures of the proteins are known to be conserved more than amino acid sequences, homology modelling generally gives accurate results [25]. Because of its high reliability homology modelling has been widely used in many studies. Homology modelling requires 5 essential steps: Searching the databases for homologues proteins, sequence alignment of proteins, main-chain model construction, side-chain model construction and fine –tuning the model by minimization [24].

For the accuracy of the model, sequence similarity between the homologue proteins and the protein of interest should at least be more than 30% [29]. Around this similarity RMSD value of the backbones of the proteins differ near 2Å [30]. Although models obtained by 30%-50% sequence similarity may be used in some applications, a reliable model is obtained when the sequence similarity is higher then 50% with an RMSD difference less than 1 [30][31].

2.4 Molecular Dynamics

Over the past decades, improvement of the power of the computers has lead enormous development of simulations of biomolecular systems. Molecular dynamics (MD) simulations are the most common techniques in the area of computer simulations of biological molecules. It has been 32 years since the first MD simulation of a protein Bovine Pancreatic Trypsin Inhibitor (BPTI) was performed [31]. Although BPTI is very small in size and the duration of the simulation was short, the study totally changed the idea which shows proteins as inflexible structures [32].

Throughout the past years, MD simulations gave many insights about the atomic interactions and fluctuations, thermodynamical properties and kinetics of the biomolecular systems [33]. With the generation of more reliable force fields, MD simulations have yielded much more accurate results. Today MD simulations of biomolecules have three main application areas [32]:

1. MD simulations are used to show how biomolecules dynamically behave.
2. MD simulations give time averaged properties of molecules. So that one may estimate the bulk properties of a system.
3. MD simulations are performed to see the accessible conformations of a protein in terms of thermodynamics.

2.5 Molecular Docking

Physiological functioning in the living organisms is achieved via intermolecular interactions between the molecules and their specific ligands. Depending on the interactions between the biomolecular pairs, different down-stream effects occur in the cell. Understanding the physical basis of the recognition of a receptor by its ligand is very important in terms of its biological, medicinal and pharmaceutical applications and usage [34]. Although today, quantity and quality of the interactions between two such molecules are measurable and predictable through experimental techniques, the time consumed is too much and the effort spent demands a lot of money [35]. Improvements in the computer technology and the calibrated force fields enabled computational docking to be a remarkable figure in macromolecular interactions studies such as computational drug design [36]. However, the results obtained via docking simulations are still not very accurate and requires a careful interpretation [37].

Resulting with the lowest binding free energy, computational docking is used as a tool in estimating prosperous receptor/ligand complexes [34]. In yielding significant estimations in ligand/receptor complexes, there are three main lines in computational docking [36]:

1. Determination of the three dimensional (3D) structures of the molecules.
2. Prediction of the binding site on the receptor molecule.
3. Estimating the binding mode.

Chapter 3

COMPUTATIONAL METHODS

3.1 Preparation of Initial Models

3.1.1 Preparation of Initial Coordinate Files

Currently, there are 3 crystal structures of JMJD2A-tudor that have been revealed: 2QQR [3] at 1.80 Å resolution, 2QQS [3] at 2.82 Å resolution and 2GFA [21] at 2.10 Å resolution. The 2QQR structure contains the tudor domain with no missing residues. For the initial coordinates B chain was selected for 2QQR. The 2QQS structure contains the tudor domain and H4K20me3 peptide together in bound form. The 2GFA structure contains the tudor domain and H3K4me3 peptide together in bound form. Both 2QQS and 2GFA structures have missing residues that have not been located in the experiments. B and D chains was selected for both 2GFA and 2QQS structures, because the number of missing residues they have are less than A and C chains. The rest of the missing residues in the tudor protein were completed by use of homology modelling. The peptide ligands were left incomplete as they are in the crystal structure. Furthermore, selenomethionine (MSE) residues in the 2QQR structure were replaced with methionine residues in the model, since MSE residues are used for labelling purposes in the crystallization experiments

Initial JMJD2A-tudor-H4K20me3 structure was also used for the dimethylated state of the same complex. Since there was not any revealed structure of JMJD2A-Tudor-H3K9me3 complex available, initial structure was modelled employing molecular docking simulations. . Parameters for trimethyllysine and dimethyllysine residues in the peptide ligand were missing and parameterization was needed. New parameters for the non-standard residue were generated using quantum mechanical techniques.

3.1.2 Homology Modelling of JMJD2A Tudor

The SWISS-MODEL homology modelling server [38] was used to carry out homology modelling procedures. According to the demanding effort of the modelling, three different modelling modes are supplied to users in the server: automated mode, alignment mode and project mode. The automated mode is designed mainly for the tasks where sequence closeness between target and the template is high. The alignment mode is designed for comparative modelling where a target sequence is provided as an initial alignment. The project mode is designed mainly for hard modelling tasks, where optimal alignment cannot be obtained via the automated and the alignment modes.

The JMJD2A-tudor domains in the 2QQS and the 2GFA structures have missing residues at both ends of the peptides. In addition, two structures have differing sequences at first 4 amino acids. For providing agreement to this conflict, the first 4 amino acids in the 2GFA structure were replaced with the first 3 amino acids in the 2QQS structure, since the 2QQR structure has the same sequence with the 2QQS. The homology modelling procedure was carried out using the alignment interface of SWISS-MODEL homology modelling server. Initial alignments of the structures were introduced in FASTA format. The 2QQR structure was used as the target sequence and the 2GFA and the 2QQS

structures were used as the template structures. B chains of three structures were selected for target and the template sequences.

3.1.3 Parameterization of Non Standard Residues

JMJD2A-tudor domains specifically bind to trimethylated H3K4, H3K9 and H4K20 and to dimethylated H4K20. Since many force field parameters are only for standard biomolecules units, new parameters for trimethyllysine and dimethyllysine residues were required for simulations. To be consistent with the parameter set of the rest of the system, an initial parameterization procedure was carried out using quantum mechanical methods.

How articles such as electrons show both particle-like and wave-like behaviour is described by quantum mechanics. Quantum mechanical calculations are performed based on the Schrödinger equation Eq (3.1), which describes the changes in quantum states of particles:

$$\left\{\frac{-h^2}{8\pi^2 m}\nabla^2 + V\right\}\Psi(\vec{r},t) = \frac{ih}{2\pi}\frac{\partial\Psi(\vec{r},t)}{\partial t} \qquad \textbf{(3.1)}$$

where h is Planck's constant, m is mass of the particle, V is potential field , Ψ is the position and time dependant wave function and ∇ is the differential operator which indicates partial differentiation with respect to x, y and z constituents:

$$\nabla = \frac{\partial}{\partial x}\hat{i} + \frac{\partial}{\partial y}\hat{j} + \frac{\partial}{\partial z}\hat{k} \qquad \textbf{(3.2)}$$

Because of the accuracy of the effective charges and the rest of the parameters, the Duan et al. force field [39] was selected for the molecular dynamics simulations. The Duan et al. force field is also named as AMBER ff03, indicating that it is an AMBER force field and was developed in 2003. Although the Duan et al. force field parameter set has retained some of the parameters from the previous AMBER force fields (Cornell et al. [40], Wang et al. [41]), it is recognized as a distinct force field instead of an extension. The main reason for this is the difference in the charge derivation of individual atoms in amino acids. In the Cornell et al. and Wang et al. force fields charge derivations have been performed by fitting the molecular electrostatic potentials of small peptides in gas phase. Because proteins naturally occur in aqueous environments, dipole moments in the protein are affected by the solvent polarization. Neglecting the solvent effect in electrostatic potential calculations leads to overestimation of dipole moments in the proteins. However, molecular electrostatic potentials for peptides were derived in condensed phase in the Duan et al. force field, and effective charges have been obtained by RESP [42] fitting procedure. Moreover, molecular electrostatic potentials in the Duan et al. force field were derived using density functional theory (DFT) with B3LYP correlation and exchange functions and ccpVTZ basis set, thus giving more accurate results. Below is the procedure of the parameterization of the non standard residues.

Initial coordinates of the non standard residues were generated as peptide fragments made up of Ace-trimethyllysine-Nme and Ace-dimethyllysine-Nme for trimethyllysine and for dimethyllysine residues respectively. For accuracy, two different conformations were used for the peptide fragments. The first one was alpha conformation where dihedral angles were Φ, Ψ = -60, -40 respectively and the second one was the beta conformation where dihedral angles were Φ, Ψ = -120, 140 respectively. All of the three dimensional coordinates were obtained using Accelrys DS Visualizer 2.0 [43].

Geometry optimization were done at the level of Restricted Hartree-Fock (RHF) theory with 6-31G* basis set. Multiplicity value and the total charge of the peptide were introduced as 0 and +1 respectively. Dihedral angles were fixed to their initial conformational states for geometry optimization. All quantum mechanical calculations were performed using Gaussian 03 [44] and input file for Gaussian 03 was generated using the Ante_R.E.D. program from R.E.D. III tools [45].

After completing geometry optimization, molecular electrostatic potential calculation and RESP fitting procedure were performed using Gaussian 03 and R.E.D. III [45] program. Chemically equivalent methyl groups were set to have the same effective charges and the total effective charges for acetyl and methyl caps were set to 0. Molecular electrostatic potential calculations were done using DFT, at the level of B3LYP theory with ccpVTZ basis set. IEFPCM was chosen as continuum solvent model and ether, which has a dielectric constant of 4, was chosen as the organic solvent which IEFPCM applied in.

Library files for the non standard residues were created using Leap which is an AMBER tool and distributed with AMBER molecular dynamics simulation package [46] . Atom types in the non standard residues were adapted from the general AMBER force field (GAFF) [41].

3.1.4 Molecular Docking

Two components are necessary for docking simulations: a searching algorithm which suffices a good examination of conformational space occupiable by the complex and a scoring function which gives a good representation of the energetics of the interaction [41]. Exploring the search space can be achieved through several ways such as molecular dynamics, simulated annealing and genetic algorithms [47]. In this thesis, being an elaborate search algorithm method, genetic algorithms are used for sampling in docking

tools. Given the conformations and orientations of the receptor bound to ligand, scoring function determines the binding affinity based on molecular mechanics force fields. In this study docking simulations were carried out in AutoDock 4.0 [48][49]. The scoring function defined in AutoDock 4.0 is a semiempirical free energy force field, which consists of following evaluations:

$$\Delta G = (V_{bound}^{L-L} - V_{unbound}^{L-L}) + (V_{bound}^{R-R} - V_{unbound}^{R-R}) + (V_{bound}^{R-L} - V_{unbound}^{R-L} + \Delta S_{conf}) \quad \textbf{(3.3)}$$

where L stands for the ligand and R stands for the receptor in the complex. In the equation, V refers to the energetics of the corresponding component and as shown in equation 3.4, V consists of a van der Waals term, a hydrogen bond term, an electrostatic term and a desolvation term, weighted with specific constants.

$$\begin{aligned} V = {} & W_{vdw} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) + W_{hbond} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) \\ & + W_{elec} \sum_{i,j} \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} + W_{sol} (S_i V_j + S_j V_i) e^{(-r_{ij}^2 / 2\sigma^2)} \end{aligned} \quad \textbf{(3.4)}$$

ΔS_{conf} , in equation 3.3 refers to the loss of entropy coming from the torsional degrees of freedom upon complex formation. Given N_{tors} stands for the number of rotable bonds, the entropic loss contribution is defined as follows:

$$\Delta S_{conf} = W_{conf} N_{tors} \quad \textbf{(3.5)}$$

JMJD2A-Tudor-H3K9me3 is the only structure without a revealed binding mode, in this study. To predict the binding mode of the complex, molecular docking simulations were performed. The structure of the receptor was taken from the initial structure of unliganded Tudor, whereas for the ligand, the G chain of the crystal structure 2Q8C [16] at 2.05 Å resolution, was used. The receptor and the ligand structures were then minimized 10000 times separately using AMBER Sander [46] in three different conditions: in vacuum, in implicit solvent and in explicit solvent. Following the minimizations, to obtain all possible binding modes, the input files for rigid and flexible docking simulations were prepared using AutoDock Tools 1.5.2 with the addition of Gastieger charges [50]. HTD-2 of JMJD2A-Tudor was selected to be accessible for the ligand, since we assume that H3K9me3 tail also recognised by the same region with the other histone ligand complexes of JMJD2A-Tudor. Finally, with the following properties, each of the docking simulations were performed for 100 runs: Lamarckian Genetic algorithm [48] as the searching algorithm, 25,000,000 number of evaluations, population size of 250, 50,000 number of generations with the rates of 0.8 and 0.02 for mutation and for crossover respectively.

Owing to the computational reasons, only one binding mode with the lowest scoring function was selected for the molecular dynamics simulations.

3.2 Molecular Dynamics Simulations

Molecular dynamics is a straightforward reliable computational technique to obtain motions of the particles in a physical system. As this study focuses on biological systems, the centre of interest was put on biomolecules as the physical system and atoms as the particles. Based on potential energy of the system, which is obtained via calibrated force fields, molecular dynamics simulations yield kinetic movements of individual atoms on the basis of Newtonian equations of motion:

$$m_\alpha \vec{r}_\alpha = -\frac{\partial}{\partial \vec{r}_\alpha} U_{tot}(r_1, r_2, ..., r_N), \qquad \alpha = 1,2,...N \tag{3.6}$$

where m_α is the mass of particle, r is the particle's position, U_{tot} is the total potential energy that is dependant on the positions of all atoms and N is the number of the atoms in the system. U_{tot} is obtained based on the potential function that is composed of following:

$$U_{tot} = U_{bond} + U_{angle} + U_{dihedrals} + U_{Coulomb} + U_{vdW} \tag{3.7}$$

The first three terms above come from the bonded interactions between atoms:

$$U_{bond} = \sum_{bonds} K_b (b - b_{eq})^2 \tag{3.8}$$

$$U_{angle} = \sum_{angle} K_\theta (\theta - \theta_{eq})^2 \tag{3.9}$$

$$U_{dihedrals} = \sum_{dihedrals} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] \tag{3.10}$$

where K stands for a constant for bonds and bond angles, b and θ are bond length and bond angle, b_{eq} and θ_{eq} are equilibrium bond length and bond angle, V_n is a force constant , φ is dihedral angle and γ is the phase angle. The last two terms in the Eq. 3.7 come from the non bonded interactions between atoms:

$$U_{Coulomb} = \sum_{i<j}\left[\frac{q_i q_j}{\varepsilon R_{ij}}\right] \quad \textbf{(3.11)}$$

$$U_{vdW} = \sum_{i<j}\left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}}\right] \quad \textbf{(3.12)}$$

where q_i and q_j are partial charges of atoms i and j, ε is the dielectric constant, A_{ij} and B_{ij} are constants differing for different atom interactions and R_{ij} is the distance between the i^{th} and j^{th} atoms. Nonbonded interactions are computed for all non bonded atom pairs using Lennard Jones potential for van der Waals interactions and Coulomb potential for electrostatic interactions.

Procedural details about the molecular dynamics simulations that were carried out in this study are stated below.

The structures were introduced to the Leap module of AMBER 10 molecular dynamics simulations package [46][51] in PDB file format. Each one of the structures that was prepared was solvated in a water box where the distance between the edges of the water box and the protein was set to 12 Å. The TIP3P atomistic water model [52] was chosen to represent the water in the system. Since the systems were not fully uncharged, Na+ and Cl- ions were added to neutralize the systems. The AMBER force field for bioorganic molecules (ff03) [39] was used to obtain parameters for all systems.

The NAMD molecular dynamics simulations program [53] was used for all of the simulations either equilibrium and minimization steps as well as the production step.

Minimization was carried out 25000 times with a conjugate gradient method implemented in the NAMD.

Periodic boundary conditions were applied for equilibration and production run periods of the systems. The SETTLE algorithm [54] was used for keeping the bond lengths fixed in water molecules with a rigid bond tolerance of 10^{-5} Å. For full electrostatic interactions the Particle Mesh Ewald (PME) [55] regime was used since the interactions in periodic boundary conditions are extravagant. For the Lennard Jones interactions, a distance of 10 Å was used as the cutoff value. Coordinates and energies were collected in every 1ps where integration times of the simulations were chosen as 2fs. Systems were gradually annealed from 10 K to 310 K in a time period of 1500 ps. When the temperatures reached 310 K, the temperature was maintained using a Langevin thermostat with a coupling coefficient of 5/ps. Langevin dynamics were turned off for hydrogen atoms in the system. Since the simulations were performed in the isobaric-isothermal ensemble (NPT), constant pressure control was applied to the systems. Maintenance of the pressure at 1.01325 bar was carried out on the basis of Langevin piston Nose-Hoover method [56][57] with a barostat oscillation time of 100 fs, a barostat damping time of 50 fs and a barostat noise temperature of 310 K. 50 ps of equilibration period were performed for each system after minimization and annealing steps.

The production simulations of the systems were performed for 25 ns long using the methods as in the equilibration period. Coordinates and energy values were collected every 1 ps throughout the simulations.

3.3 Binding Free Energy Calculations

The non covalent association of a receptor molecule and a ligand molecule in a solution is as follow:

$$[R]_{aq} + [L]_{aq} \Leftrightarrow [C]_{aq} \quad \textbf{(3.13)}$$

where R stands for receptor, L stands for ligand and C stands for the complex that the receptor and the ligand form together. The association of the molecules generates a free energy difference that is related to the binding free energy of the ligand. The experimental binding free energy of the ligand is computed from experimental studies as follow:

$$\Delta G = RT \ln(K_d) \quad \textbf{(3.14)}$$

where R being the molar gas constant, T being the temperature and K_d being the dissociation constant obtained experimentally. The binding free energy of the ligand is also computed as the free energy difference between complex and receptor and ligand:

$$\Delta G_{binding} = G_{Complex} - G_{\mathrm{Re}ceptor} - G_{Ligand} \quad \textbf{(3.15)}$$

The free energy is formed of enthalpic end entropic contributions:

$$G = H - TS \quad \textbf{(3.16)}$$

where H is the enthalpy, T is the temperature and S is the entropy of the molecule. The Enthalpy of each of the molecules given in Eq. 3.16 is composed of two components: solute effect to free energy and the solvent effect to free energy. To see the solute and solvent contributions, free energy may be restated:

$$G = E_{MM} - TS_{MM} + G_{solvent} \quad \textbf{(3.17)}$$

where the first term E_{MM} is the average energy of the solute and comes from the bonded and non bonded molecular mechanics interactions:

$$E_{MM} = E_{Bond} + E_{Angle} + E_{Torsional} + E_{vdW} + E_{Coulomb} \quad \textbf{(3.18)}$$

where E_{bond} E_{Angle} and $E_{Torsional}$ contributions stand for bonded interactions and E_{vdW} and $E_{Coulomb}$ stand for non bonded interactions. In a computer simulation these contributions are obtained from molecular dynamics simulations.

The second term TS_{MM} in Eq 3.17 comes from the entropic contribution of the solute. T represents temperature and S_{MM} represents the entropy that is obtained from molecular mechanics. In detail S_{MM} is consisted of following terms:

$$S_{MM} = S_{Rotational} + S_{Translational} + S_{Vibrational} + S_{Configurational} \quad \textbf{(3.19)}$$

where $S_{Rotational}$, $S_{Translational}$, $S_{Vibrational}$ and $S_{Configurational}$ stand for rotational, translational, vibrational and configurational motions of the solute respectively. It should be noted that, the values coming from $S_{Configurational}$ terms were neglected in this study.

The last term $G_{solvent}$ in Eq. 3.17 comes from the solvent contribution to free energy is composed of two components:

$$G_{Solvent} = G_{Polar} + G_{Nonpolar} \quad \textbf{(3.20)}$$

where G_{Polar} stand for the polar contribution and is computed via Poisson-Boltzmann (PB) and Generalized Born (GB) methods in this study. The second term $G_{Nonpolar}$ stands for the nonpolar contribution and is computed from solvent accessible surface area (SASA):

$$G_{Nonpolar} = \gamma SASA + \beta \qquad \textbf{(3.21)}$$

where γ stands for surface tension, SASA stands for the solvent accessible surface area of the solute and β stands for an offset value.

3.4 MM-PBSA/GBSA and Normal Mode Calculations

PB calculations were carried out in DelPhi: a finite difference Poisson-Boltzmann solver program [58] using 2400 snapshots from the last 24 ns of the molecular dynamics simulations with 10 ps time intervals. Three PB calculations were performed for each structure based on different internal dielectric constants 1, 2 and 4 for protein and one external dielectric constant 80 for solvent. Parse radii [59] and Duan et al. charges were employed and the modified Bondi radii [60] were augmented by 1.4Å for PB calculations. For the SASA calculations the Molsurf program [61], which is a part of AMBER simulation package, was used with the LCPO method [62]. γ value was taken as 0.00542 and β value was taken as 0.92 for nonpolar contributions.

To compare the results of PB calculations another method, GB calculations were performed. GB calculations were carried out in the AMBER MM-PBSA tool using the GB solver. The modified GB model, which was proposed by Onufriev at al. [60], was selected for calculations. Three GB calculations were performed as in the PB calculations. The SASA calculations were performed the same as reported above. In GB calculations γ value was taken as 0.005 and β value was taken as 0 for nonpolar contributions.

To find the hot-spots of the protein, free energy decomposition calculations were also performed in the AMBER MM-PBSA tool using the GB model. As in the PB calculations three decomposition calculations were performed for each structure. γ and β values were taken the same as in the GB calculations. Since the decomposition calculations work only with the ICOSA method, for the SASA calculations the ICOSA method was utilized.

Normal mode calculations were carried out in the AMBER NMODE module to find the entropic contributions of association. Because of the computational expense of the NMODE calculations, 240 snapshots were used from the last 24 ns of the molecular dynamics simulations with 100 ps time intervals. The calculations were performed using a distance dependant dielectric constant $\varepsilon = 4R_{ij}$, that was applied for each structure.

Table 3.1: Setup parameters for the systems used in molecular dynamics simulations.

System	Initial Structure	Parameterized Residue	Homology Modelling	Docking	MD
Unliganded JMJD2A Tudor	2QQR				25 ns
JMJD2A Tudor-H3K4me3	2GFA	Trimethyllysine	Yes		25 ns
JMJD2A Tudor-H4K20me3	2QQS	Trimethyllysine	Yes		25 ns
JMJD2A Tudor-H4K20me2	2QQS	Dimethyllysine	Yes		25 ns
JMJD2A Tudor-H3K9me3	2QQR	Trimethyllysine		Yes	25 ns
H3K4me3	2GFA	Trimethyllysine			25 ns
H4K20me3	2QQS	Trimethyllysine			25 ns
H4K20me2	2QQS	Dimethyllysine			25 ns
H3K9me3	2Q8C	Trimethyllysine			25 ns

Chapter 4

CONVERGENCE AND STABILITY

4.1 Convergence of Trajectories

To assess the convergence of the trajectories, autocorrelation functions were calculated to obtain autocorrelation times. The formula for the autocorrelation function is defined as:

$$R(k) = \frac{\sum_{i=1}^{N-k}(X_i - \overline{X})(X_{i+k} - \overline{X})}{\sum_{i=1}^{N}(X_i - \overline{X})^2} \qquad \textbf{(4.1)}$$

where X_i stands for the measurement at time step i, $\overline{X}$ stands for the mean of all measured X values and k stands for the lag. For different lags a time dependant function was obtained and autocorrelation time was obtained by fitting this function to the following exponential model:

$$R(t) \sim \exp(-t/\tau) \qquad \textbf{(4.2)}$$

where t stands for time and τ stands for the autocorrelation time.

Table 4.1: Autocorrelation time coefficients of the curves fitted to the autocorrelation functions with 95% confidence bounds.

Structure	τ **(ps)**
Unliganded JMJD2A Tudor	2.23
JMJD2A Tudor-H3K4me3	2.25
JMJD2A Tudor-H4K20me3	2.28
JMJD2A Tudor-H4K20me2	2.26
JMJD2A Tudor-H3K9me3	2.23
H3K4me3	2.17
H4K20me3	2.18
H4K20me2	2.24
H3K9me3	2.19

To obtain autocorrelation times, different measured quantities were used such as RMSD, radius of gyration and energy components of the systems. Since RMSD and radius of gyration values were highly fluctuating, total energy values of the systems were used as the reference data. Resulting autocorrelation times, which were obtained using the function in Eq. 4.2, are demonstrated in Table 4.1.

Based on the backbone atoms, RMSD values were calculated in each system. Because of the globular motions that were observed during the simulations, as well as RMSD values of the overall structures, RMSD values of each of the HTD and the histone tail calculated separately for each system (Figure 4.1).

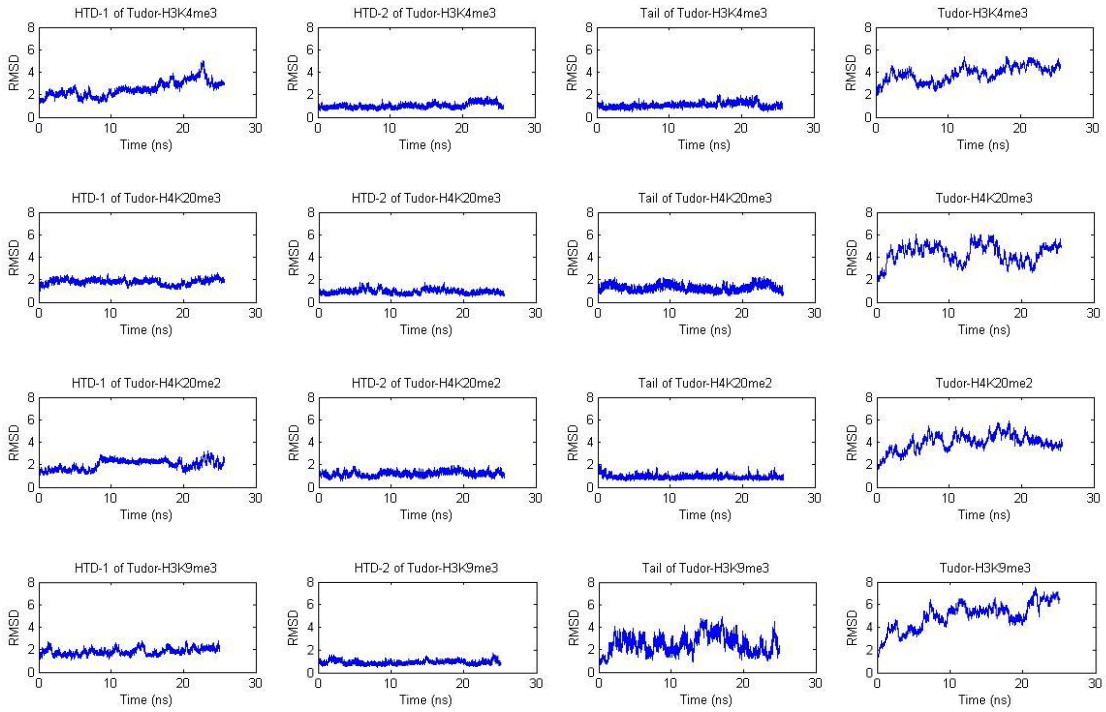


Figure 4.1: Root mean square deviations (RMSDs) for the complexed structures versus time. RMSDs were computed for each of the Hybrid Tudor Domain 1 (HTD-1), Hybrid Tudor Domain 2 (HTD-2) and histone tail along with the overall structures.

As seen in figure 4.1, although HTD-2 looks equilibrated in all structures, RMSD values of HTD-1 still fluctuate. Main reason behind this deviation leans to unrestrained terminal residues which move freely. Removing the several terminal residues from the RMSD calculations reduces the RMSD values significantly. Of the four complex structures, the histone tail in JMJD2A-Tudor-H3K9me3 appears to be the most fluctuating one with respect to the other tails. This is mostly due to the relatively less favourable binding of H3K9me3. Furthermore, it is noteworthy to remember that this structure was initially modelled; therefore, it is not surprising to observe high RMSD values. In this

sense, high deviations of the RMSD values of all of the structures are reasonable and are not confusing.

To better understand the effects of globular motions to RMSD values, radius of gyration values were calculated for each structure. The radius of gyration of a molecule is defined as:

$$r_g^2 = \frac{\left(\sum_{i=1}^{n} m_i (r_i - \bar{r})^2\right)}{\left(\sum_{i=1}^{n} m_i\right)} \qquad \textbf{(4.3)}$$

where m_i is the mass of the i^{th} atom, r_i is the position of the i^{th} atom and $\bar{r}$ is the weighted centre of the molecule. Since radius of gyration is a measure of the shape of a molecule, it gives specific information about the characteristics of a molecule. More discussion on the radius of gyration analysis together with the flexibility of the structures is available in section 5.2.

4.2 Time Series Convergence

As discussed previously in Methods section, the MM-PBSA/GBSA and the NMODE methods were employed for the calculation of the binding free energies. To obtain internally consistent results, values obtained from the MM-PBSA/GBSA and the NMODE methods should be subjected to the convergence analysis. Since the MM-PBSA/GBSA and the NMODE methods require snapshots from an equilibrated trajectory, first 1 ns of simulations were removed. Although initial structures were obtained from the experimental crystal structures of the proteins, taking into account the high deviations in the protein conformations involved a long equilibration time period. Considering the nonessential

modes to be eliminated from the protein dynamics, 1 ns of time was a reasonable period to obtain meaningful results. From the rest 24 ns of simulations, 2400 snapshots for each structure and for each of the three species were obtained with 10 ps equal time space. Since taking small time intervals between each snapshot may be inaccurate due to the motional correlations, 10 ps of spacing was chosen to be a good time interval although autocorrelation times that were obtained from the total system energies were no more than 3 ps.

To determine the convergence of enthalpic and entropic contributions to the binding free energy, mean and the standard error were calculated for each structure. Considering the length of the simulations, with a 10 ps of time interval between each uncorrelated snapshot, it was not unexpected to obtain the values of enthalpy with a high deviation. As can be seen in the RMSD and the radius of gyration figures (Figure 4.1 and Figure 5.5), the protein undergoes large conformational changes that lead large differences in the energy calculations throughout the trajectories. It should be noted that, three trajectory approach was carried out in this study. In three trajectory approach, snapshots from each of the three species (complex, receptor and ligand) are collected separately that might result in large delta values, since each of the trajectories are independent. Energy fluctuations in different time periods might lead to big deviations from the mean values in three trajectory approach. Moreover, the MM-PBSA/GBSA tools represent solvent effect implicitly in a corresponding continuum solvent approach that would not always suffice the accuracy criteria.

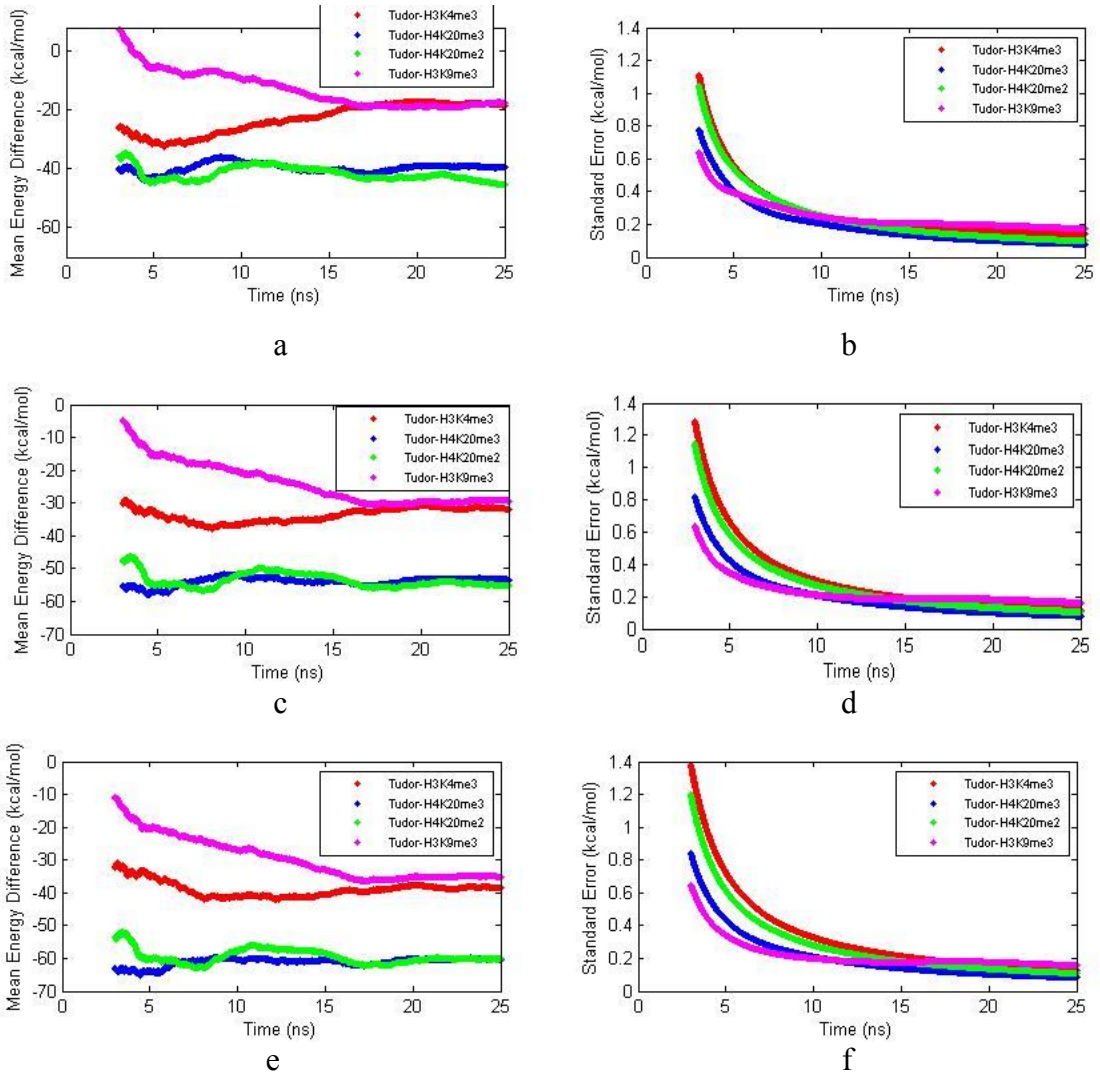


Figure 4.2: Convergence of the mean values of the GB enthalpies for the internal dielectric constant of 1 (a), 2 (c) and 4 (e) and convergence of the standard errors of the GB enthalpies for the internal dielectric constant of 1 (b), 2 (d) and 4 (f).

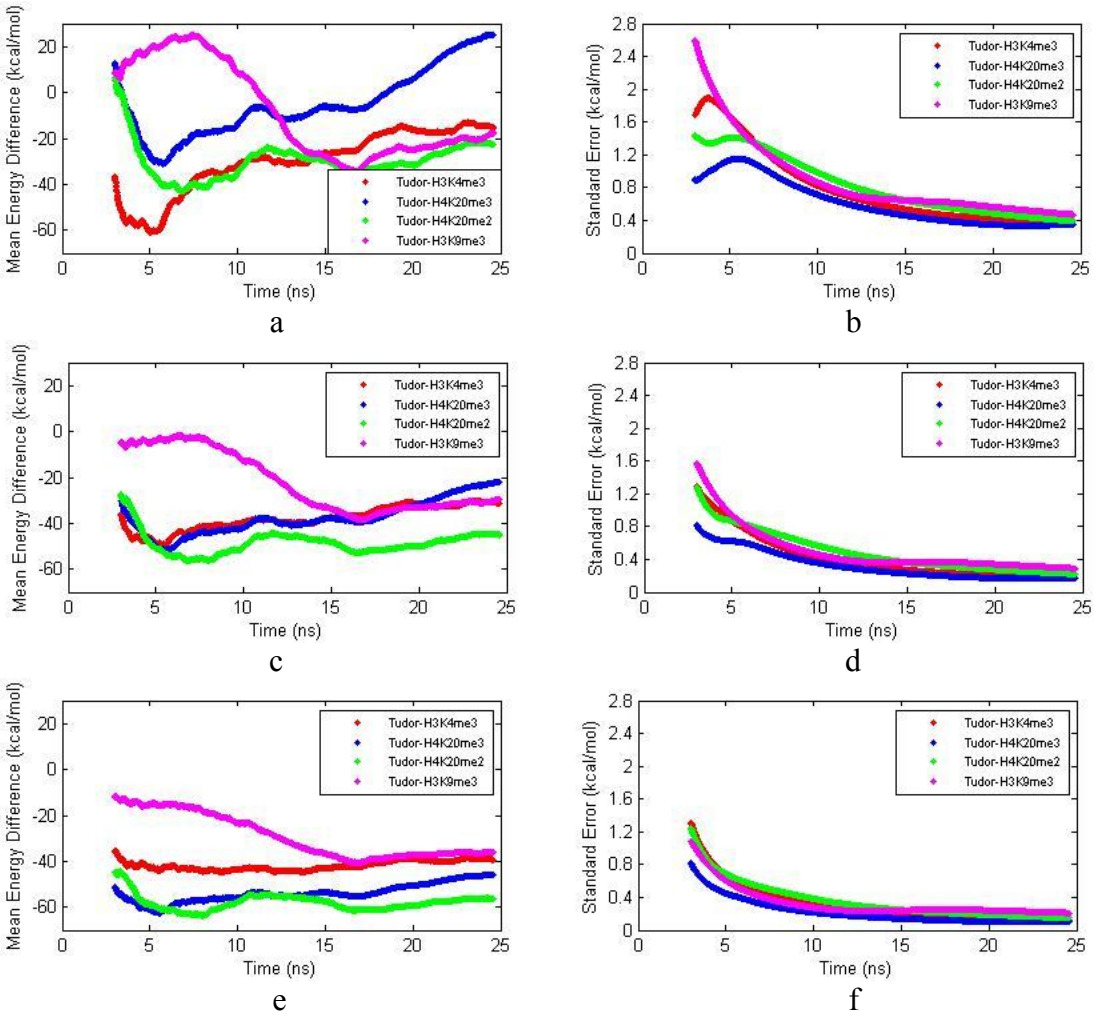
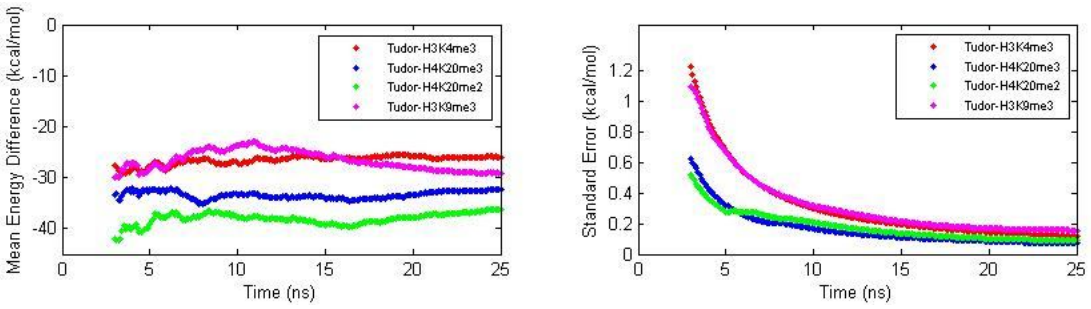


Figure 4.3: Convergence of the mean values of the PB enthalpies for the internal dielectric constant of 1 (a), 2 (c) and 4 (e) and convergence of the standard errors of the PB enthalpies for the internal dielectric constant of 1 (b), 2 (d) and 4 (f).

As expected, GB calculations were the least computational power demanding ones, whereas NMODE calculations were the most demanding ones. PB approach was also

employed in this study to compare the values of enthalpic contribution to the binding free energy.

Since there is not any common knowledge about the best representation of the internal dielectric constant of the proteins, in these binding free energy calculation methods, three different internal dielectric constants: 1, 2 and 4 were used for each approach. Convergence of the enthalpy values in GB approach show that the mean value of enthalpy converges fast when the internal dielectric constant is 1. However, ΔH values converge fast when the internal dielectric constant is 4 in PB approach. As can be seen from the figures 4.2 and 4.3, enthalpic contributions obtained from the PB calculations are less convergent than the enthalpic contributions obtained from the GB calculations. Entropic contributions also satisfy convergence criteria as can be seen in the figure 4.4.



a b

Figure 4.4: Convergence of entropic contributions to binding free energy by means of the mean (a) and the standard error (b).

Chapter 5

STRUCTURAL ANALYSIS

5.1 Binding Site Differences

In this study, we beheld that, there were two different binding modes for the recognition of the H3K4me3, H4K20me3, H4K20me2 and H3K9me3 peptides. The two different binding modes (Figure 5.1) were distinguished by the orientations of the peptides which were located in the opposite directions relative to each other and their interactions with different residues located on the protein. The first binding mode was adopted by the H3 peptides, whereas the other mode was adopted by the H4 peptides. Since the starting structures were same but the methylated lysine residues, not surprisingly, H4K20me3 and H4K20me2 adopted the same binding mode. In agreement with the experimental data; on the other hand, H3K4me3 adopted a different binding mode together with H3K9me3. Having compared with each other, interactions between the receptor molecule and the H3 peptides were related; thus revealing that H3K9me3 is recognized by JMJD2A-tudor with a similar fashion.

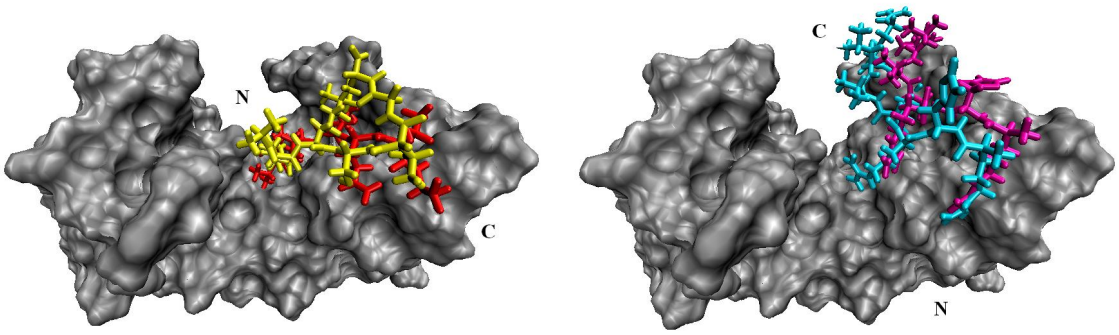


Figure 5.1: Two different binding modes of JMJD2A-tudor (A) with liganded to H3K4me3 (red) and H3K9me3 (yellow) (B) with liganded to H4K20me3 (cyan) and H4K20me2 (magenta). First two peptides bind to the HTD-2 with a similar mode in the same orientation, whereas the later peptides bind in the opposite orientation with a similar mode.

Consistent with the experimental data, methylated lysine residues of the peptides were caged by the aromatic side chains of Phe932, Trp967 and Tyr973 of JMJD2A-tudor HTD-2 throughout the simulations. In this aromatic cage, methyl groups of the trimethyllysine residues were rotating freely, whereas the methyl groups of the dimethyllysine residue was stable during the molecular dynamics simulations. To understand this behaviour, time evolution of the flexibility of the methylated lysine residues were investigated by means of torsional angles defined by C_δ, C_ε, N_ζ and CZ atoms. As illustrated in figure 5.2, trimethylated structures display similar behaviour in terms of flexibility of the methylated residues. Conversely, methyl groups of the dimethylated structure show a distinct fluctuation pattern. The defined torsional angles of the trimethylated residues were mostly oscillating in between ~ 40 to 80 degrees, -40 to -80 degrees and -160 to 160 degrees; whereas the angles of the dimethylated lysine were mostly oscillating in between ~ 50 to 90 degrees and -150 to 170 degrees. This observation indicates that the methyl groups in the methylated lysine residue, depending on the number of methyl groups, retain two or three

conformations with high fluctuations. More intriguingly, conformations of these subgroups shifted continuously by rotating throughout the simulation for the trimethylated lysine residues; while the methyl groups in the dimethylated lysine residue retained their conformations throughout the simulation. This finding suggests that, trimethyl groups are less stable with respect to the dimethyl group, in terms of their dynamic fluctuations. To seek the underlying reason for the varying stability, the overall changes in the neighbourhood of the methylated residues should be investigated carefully. Absence of a methyl group in the binding pocket of the dimethylated lysine residue, leads to strengthening of the intermolecular interactions of the residues in the vicinity of the binding pocket. Hence, the methyl groups of the trimethylated lysine residues shift their conformations; whereas the methyl groups might be localized steadily due to the high energetic barrier for the dimethylated lysine residues.

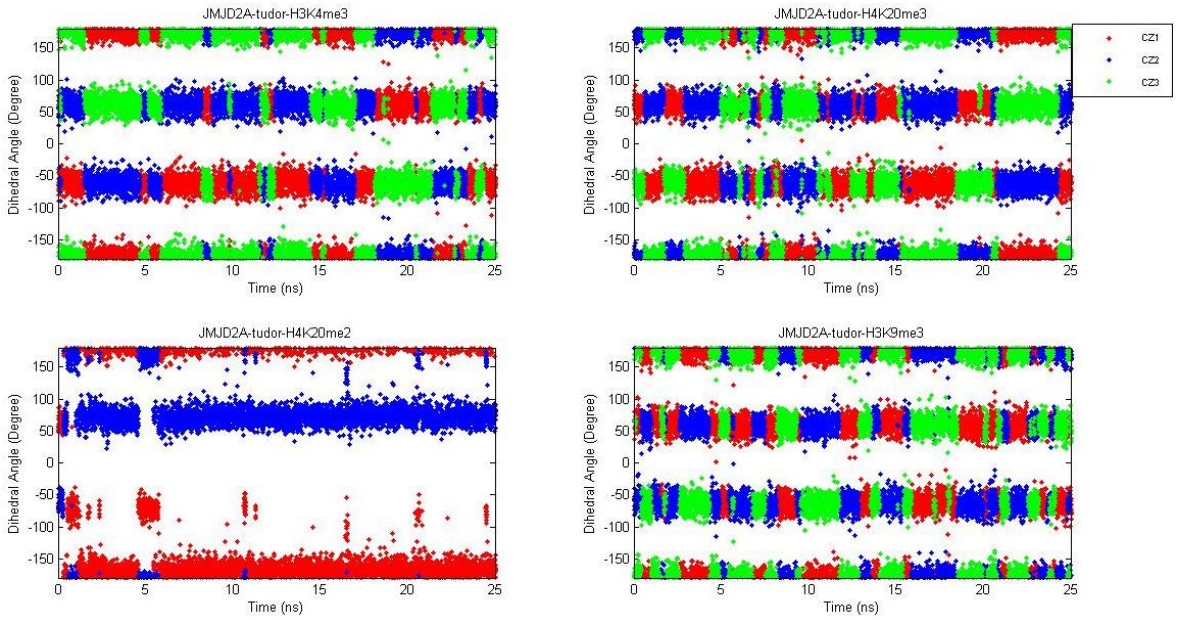


Figure 5.2: Dihedral angles of the methylated lysine residues defined by C_δ, C_ε, N_ζ and CZ atoms in the Tudor molecule liganded to H3K4me3, H4K20me3, H4K20me and H3K9me3 respectively.

5.2 Flexibility of the Complexes

In this thesis, flexibility of the structures was investigated by computing the root mean square fluctuations (RMSF) and radius of gyration values carefully. Root mean square fluctuations of backbone atoms were calculated and mass averaged for each residue after carrying out backbone RMS fitting to the initial structure. As illustrated in figure 5.3, RMSFs versus residue numbers are given for all of the complex structures and the free receptor structure.

The average RMSF values for the Tudor-H3K4me3/H4K20me3/H4K20me2/H3K9me3 complexes and the free tudor itself are 1.53, 1.29, 1.41, 1.45 and 1.50 Angstroms

respectively. As shown in the figure 5.3, all of the structures share a similar RMSF distribution with small variations. Of the five structures, the Tudor-H3K4me3 complex is the most mobile, whereas the Tudor-H4K20me3 complex is the least mobile with respect to the other structures. Relative high mobility of the Tudor-H3K4me3 structure is because of the freely moving terminal residues and it does not appear to be correlated with ligand binding.

The region between the residues 940-965 is very stable with respect to the other regions in all of the five complexes. Despite the fact that this region has not observed to be involved in binding to ligand in any of the structures, the distribution pattern of RMSF values around this region should cautiously be investigated. It is seen that the tudor-H4K20me3 structure has relatively lower RMSF values around this region. The low calculated binding free energy for this structure may also be suggestive for the low fluctuations compared to the free Tudor structure. Another observation is that upon binding to ligand, the flap regions between the β1' and the β2' strands and the β3 and the β4 strands at the binding site of the receptor (residues 930-940 and 965-973) display less displacement as expected. The fluctuations of the corresponding regions at the HTD-1 (residues 907-915 and 988-998); on the other hand, do not result in remarkable changes upon complex formation.

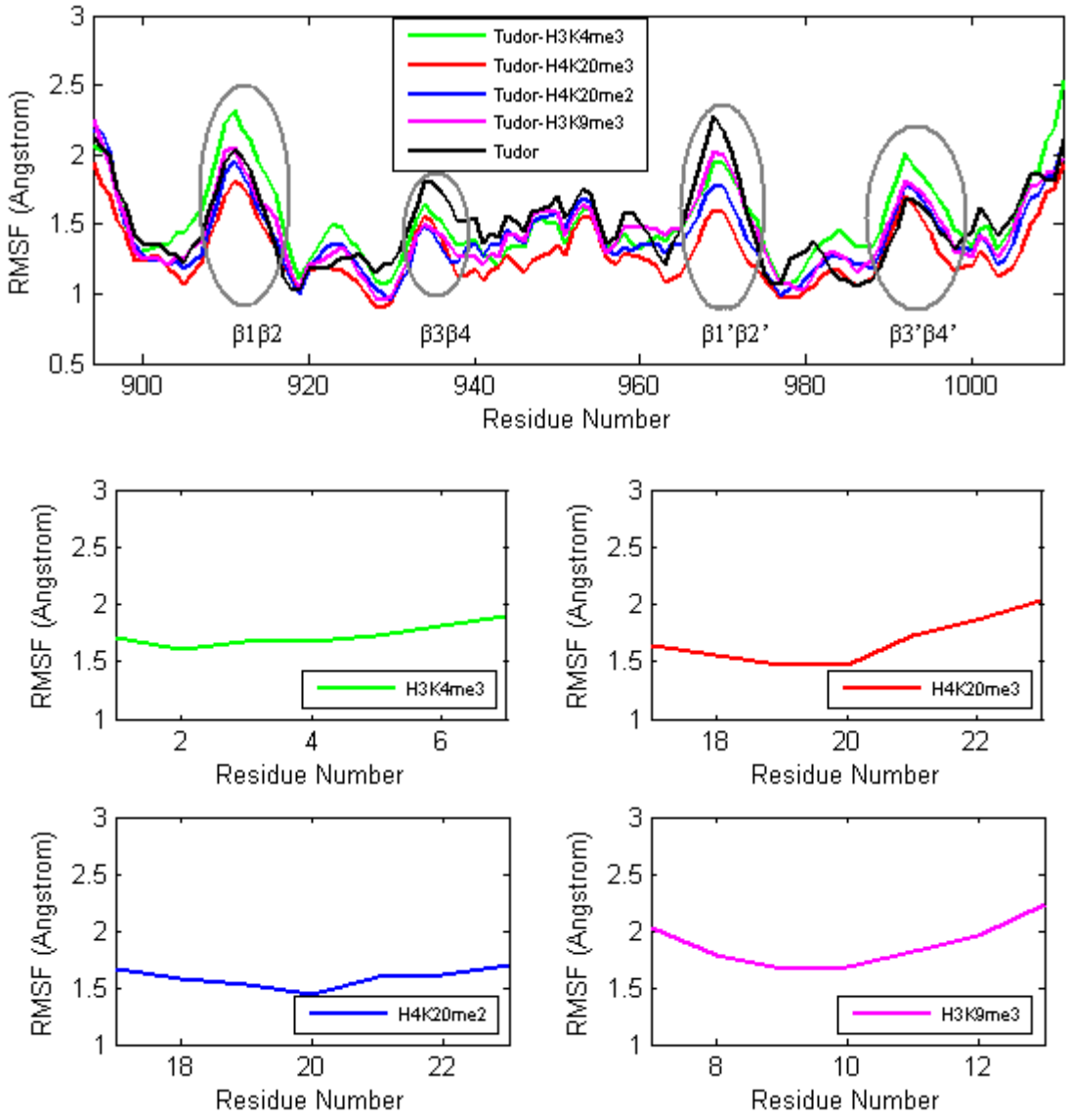


Figure 5.3: Root mean square fluctuations of the C, N and Cα atoms of the complex structures and the unliganded structure versus residue number in the structure.

The RMSF values are 1.72, 1.68, 1.58 and 1.87 for the H3K4me3, H4K20me3, H4K20me2 and H3K9me3 ligands in the complex structures respectively. In this respect, H4K20me2 tail is observed to be the most stable ligand, whereas the H4K20me3 tail is the least stable. Furthermore, methylated lysine residue in H4K20me2 is the least mobile residue relatively to the methylated lysine residues in other structures; thus indicating strong interactions with neighbouring residues. This may imply that H4K20me2 ligand is more favourable upon binding to the receptor compared to H4K20me3 ligand.

In this study, it was found that during the molecular dynamics simulations, the β1β2 and β1'β2' flap regions were very mobile and the tips of these regions were found to be moving towards and against each other periodically (Figure 5.4). Moreover, it was seen that the periodicity was varying between each of the structures. The behaviour of these positional fluctuations was inquired further by investigating the global motions of the structures.

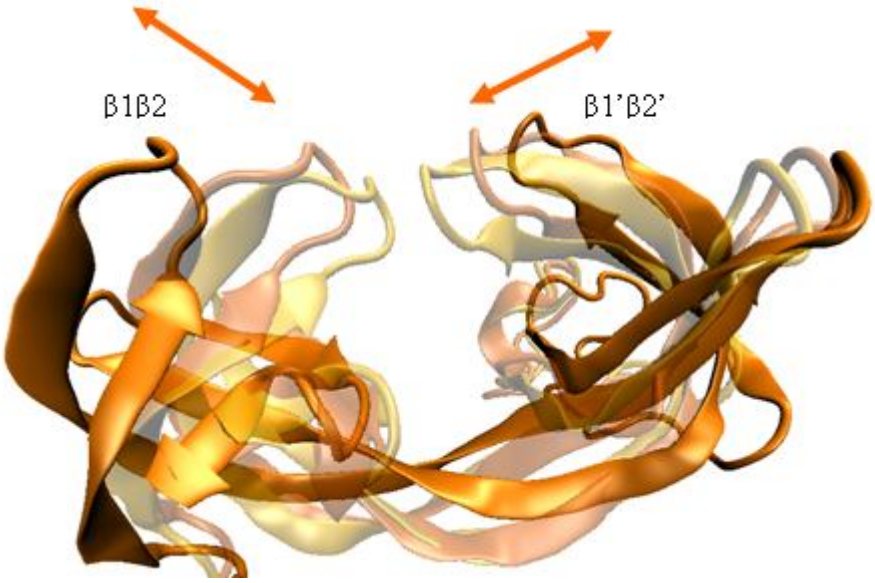


Figure 5.4: Motion between the tandem hybrid tudor domains of JMJD2A.

Determined by the change in the radius of gyration values, the Tudor domains show a periodic change in structural shape. This behaviour is caused by harmonic oscillations in global motions of the protein. It was observed that, in each structure HTD-1 and HTD-2 parts periodically come closer to each other and move apart from each other. This motion was highly dominant in the structure where there is no bound histone tail. It was observed that when the protein is bound to the histone tail, the change in the structural shape has a small frequency proposing that binding has an important role in global motion of the protein. As the simulation times get longer, radius of gyrations get smaller in bound structures, indicating that when histone tail binds to HTD-2 lobe, HTD-1 lobe, interplaying with the binding site, is undergone cooperative motions with HTD-2.

Crystal structures of the Tudor domains complexed with H3K4me3 and H4K20me3 propose distinct binding structures. As shown in figure 5.5, although the radius of gyration values are initially around 18, the values get smaller as the simulation time goes. It has been previously suggested that free and bound state conformations do not differ from each other. However, dynamics and conformations showed a distinct change in free and bound states in this study.

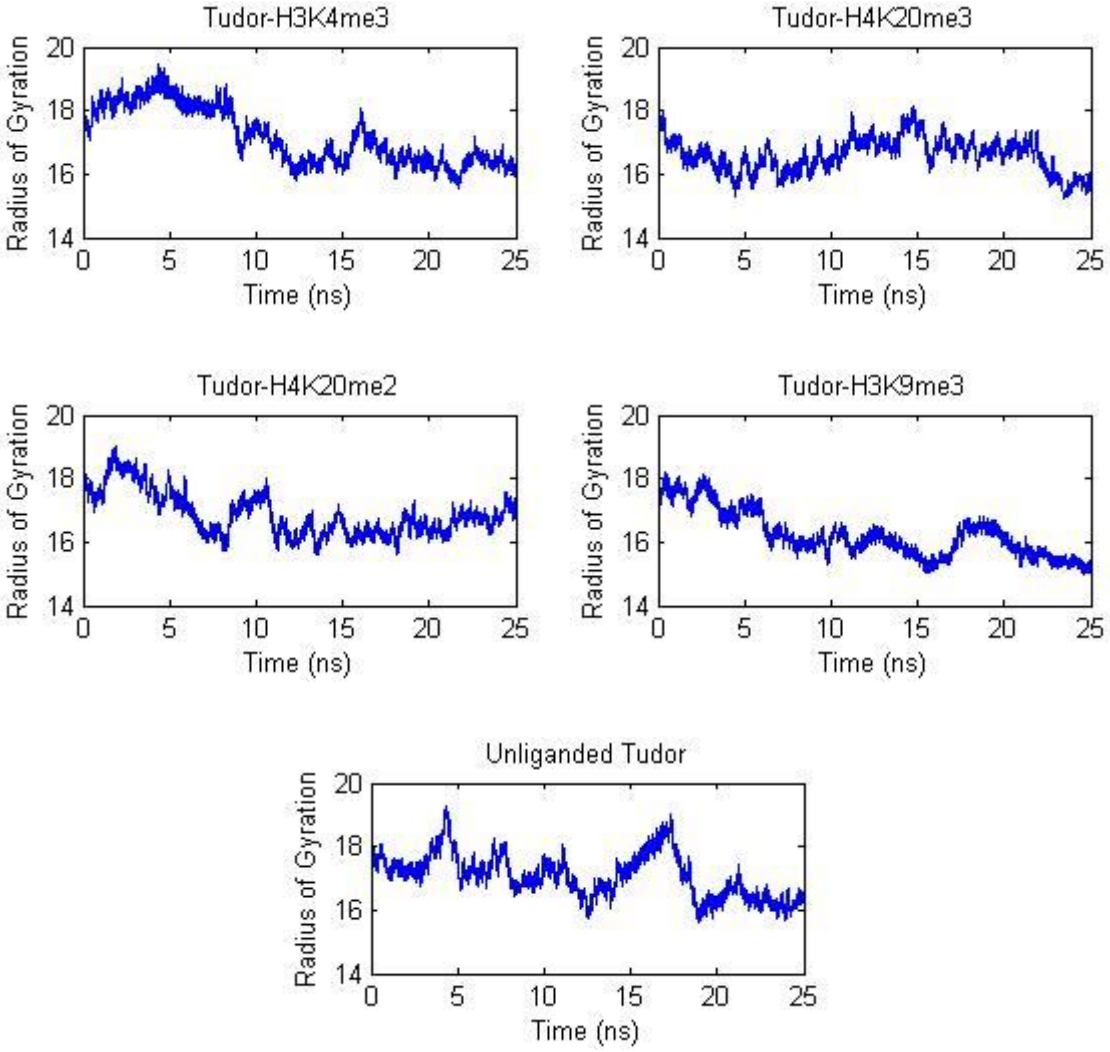


Figure 5.5: Radius of gyration of the proteins versus time for each complex structure and for the receptor structure.

5.3 Hydrogen Bonds and Salt Bridges between JMJD2A-tudor and the ligands

The molecular dynamics of the JMJD2A-tudor with H3K4me3 and with H4K20me3 structures are in a good agreement with the experimental studies. The important hydrogen bonds and the important electrostatic interactions in ligand binding were observed throughout the trajectory with high occupancies (see Appendix B and C). Although a simulation time of 25 ns may lead protein to undergo wrong conformations or interactions, in both of the structures all the interactions between the ligands and the receptors seemed very reasonable.

5.3.1 JMJD2A Tudor with H3K4me3

In the JMJD2A-tudor binding to H3K4me3 structure, Asp945 was reported to be one of the most important residues in ligand binding by interacting with Arg2 of the histone tail. Indeed, Asp945 interacted with Arg2 of histone tail with a high occupancy in molecular dynamics simulation (Figure 5.6). While forming a hydrogen bond via its HH21 atom on NH2 atom, Arg2 also formed another hydrogen bond via its HH22 atom on NH2 atom with Glu944 with a less occupation rate. In experimental studies, Asn940 was also found to be very important in binding to histone tail. Asn940 interacted with Thr3 of histone tail with occupancy of 12.78% throughout the simulation. Another important interaction was observed between Thr3 of histone tail and Ser938 as shown in figure 5.6. Although previously reported, in molecular dynamics studies Ser938 formed hydrogen bonds with the backbone amide of Thr3 with occupancy of 97.69% during the simulation. Supporting the experimental data, Tyr942 and Thr968 did not form any significant interactions with the methylated histone tail. Gln5 of the histone residue was observed to be interacting with Asp939 in the first 2.5 ns and between 7.5-15 ns of the simulation. Although, it has been

proposed that, Asp 939 had found not to be involved in any interaction with a residue on H3K4me3 histone tail, in these time intervals a hydrogen bond was formed with 11.86% occupancy. For the rest of the simulation Gln5 formed weak hydrogen bonds with the amide group and the terminal oxygen atom of H3K4me3 Ala7. Additionally, Asp 939 formed strong hydrogen bonds with His981 and Asn940 depicting it is not involved in a significant interaction with the ligand.

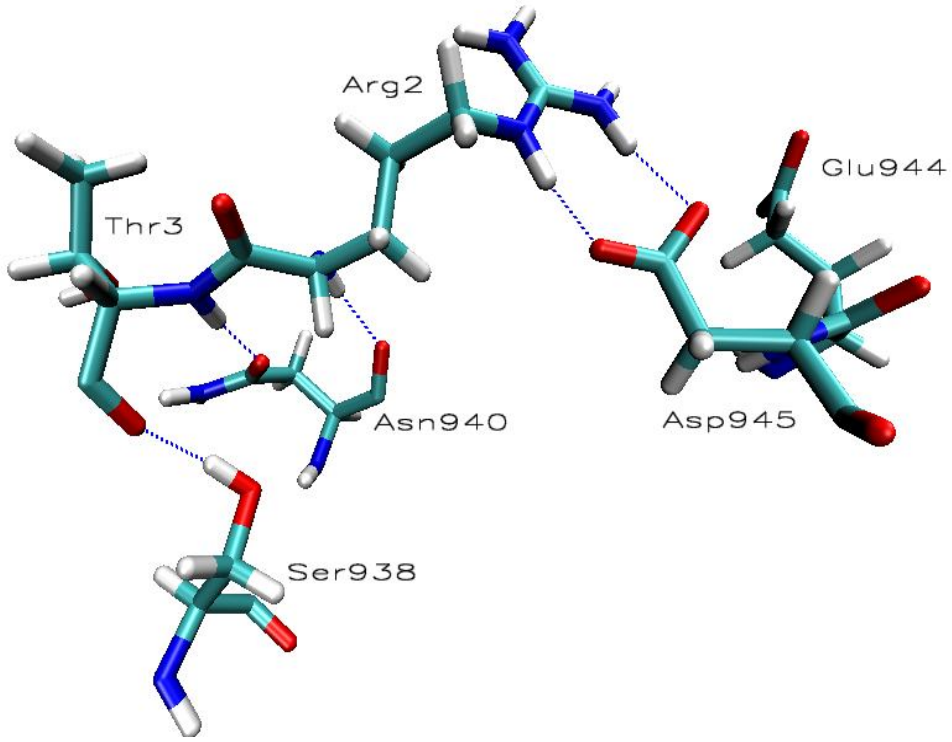


Figure 5.6: Critical residues involved in forming hydrogen bonds and salt bridges between JMJD2A-tudor and H3K4me3 peptide. Hydrogen bonds are shown in dashed blue lines.

In this study OD1 and OD2 atoms of Asp945 formed salt bridges with NH2 and NE atoms of H3K4me3 Arg2 with simulation time length of lifetime. Arg2 was also formed

important coulombic interactions with OE atoms of Glu944. NH1 and NH2 atoms of Arg2 were involved in the second type interactions. The last remarkable interaction was observed between Asp934 and H3K4me3 Lys4. Since this last interaction was not continuous during the simulation and had a less occupation rate, it may not have an important role in binding.

5.3.2 JMJD2A Tudor with H4K20me3

In the JMJD2A-tudor binding to H4K20me3 structure, Asp939 was proposed to be highly important in the binding of JMJD2A-tudor domain to H4K20me3 peptide. In the molecular dynamics simulation, Asp939 was observed to form hydrogen bonds with H4K20me3 Arg19 as it is expected (Figure 5.7). As expected from the knowledge of previous studies, Tyr942 was not involved in any hydrogen bond interactions whereas Thr968 formed some hydrogen bond interactions with the terminal residue of the histone tail.

A remarkably strong interaction occurred between Glu929 and Arg17 in the last 20 ns of the MD simulations (Figure 5.7). With a high occupancy, HH12 atom on NH1 atom and HH22 atom on NH2 atom of Arg17 were hydrogen bonded to OE1 and OE2 atoms of Glu929. Although weak hydrogen bonds were formed with Arg19 and with the backbone amide of Phe932 and the backbone amide of Ser936, they appeared to be short time interactions. Although there is not any available mutation study on this residue, the strong hydrogen bonds suggest that Glu929 might have a significant role in JMJD2A-tudor to H4K20me3 binding. Another strong hydrogen bond interaction appeared between the backbone oxygen atom of Leu922 and HE1 atom on NE1 of Trp967. This hydrogen bond had 35.16% occupancy and was consistent throughout the simulation. Trp967 also formed another hydrogen bond with the terminal oxygen and the backbone oxygen atoms of H4K20me3 Arg23 with a less occupancy compared to the first one. The residues that were

observed to be highly important in binding to H3K4me3 (Asn940 and Asp 945), had no effect on H4K20me3 ligand. Asn940 was not involved in any hydrogen bond interaction, whereas Asp 945 was involved in strong hydrogen bonds with Arg966.

In a good agreement with the previous experimental studies, Asp939 was observed to be salt bridged to Arg19 of the histone tail. Another coulombic interaction was formed between Glu929 and Arg17 consistently. The last remarkable coulombic interaction between the receptor and the ligand was formed between Asp969 and Arg23 in the last 15 ns of the simulations with a limited occupancy. Asp 945 was also observed to be salt bridged to Arg966.

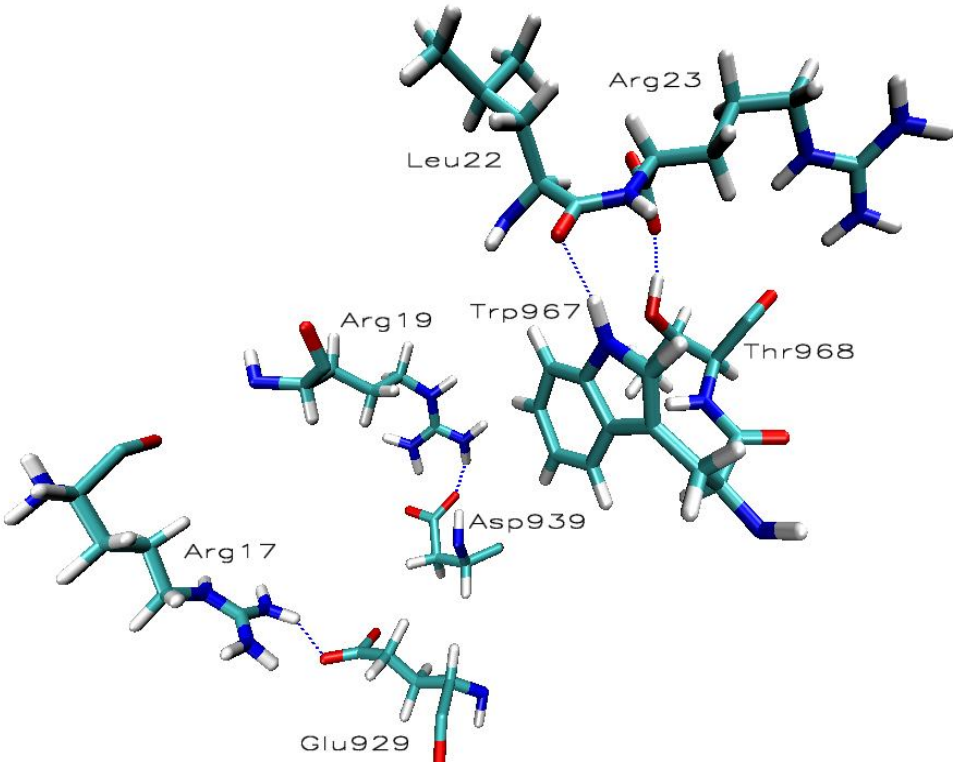


Figure 5.7: Critical residues involved in forming hydrogen bonds and salt bridges between JMJD2A-tudor and H4K20me3 peptide. Hydrogen bonds are shown in dashed blue lines.

5.3.3 JMJD2A Tudor with H4K20me2

As discussed earlier, JMJD2A-tudor domain also binds two other methylated histone tail structures. The first one is H4K20me2 that only differs from the previous structure by its dimethylation state on Lys20 and the second structure is H3K9me3. Since there is not any experimental data available yet, all interactions revealed from the molecular dynamics simulations are very important and may have a prominent role in understanding the tudor domain.

In JMJD2A-tudor-H4K20me3 structure, Arg17 of histone tail was found to have important interactions with Glu929 whereas in JMJD2A-tudor-H4K20me2 structures, Arg17 was observed to interact with two other residues. Arg17 formed Hydrogen bonds with the backbone oxygen atom of Phe932 and the backbone oxygen atom of the Ser936. Arg17 interacts with these residues in the first 12 ns of the simulation with high occupancies. Hydrogen bond interactions with Glu929 came out within the last 10 ns of the simulation proposing that Glu929 may not have much significance in binding to H4K20me2. However, this may also has caused by the starting structure of the simulation. Like the tudor binding to H4K20me3 structure, Trp967 was observed to interact with the backbone oxygen atom of H4K20me2 Arg23 via a strong hydrogen bond (Figure 5.8). Unlike the previous structure, this bond was perennial and had a high occupation of 76.90%. Arg23 also formed another hydrogen bond with OD1 atom of Asp969, in the last 15 ns of the simulation. H4K20me2 Arg19 formed unstable hydrogen bonds with Asp939 suggesting that H4K20me2 Arg19 have less importance than H4K20me3 Arg19. The dimethylated Lys20 residue was also involved in forming some hydrogen bonds with its side chain. The backbone oxygen atom of Asp 934 formed weak hydrogen bonds with the dimethylated lysine residue. Leu22 of histone tail was another residue that formed hydrogen bonds with the dimethylated lysine residue. Compared to the other, this second

interaction was continuous throughout the simulation with a high occupation. Thr968, Asn940 and Asp945 were not occupied in any interaction with the histone tail ligand. As observed in the previous structure Asp945 formed a strong hydrogen bond with Arg966.

In the molecular dynamics simulation strong coulombic interactions were formed between H4K20me2 Arg19 and Asp939. Arg23 of the histone tail and Asp969 were also salt bridged to each other. The last remarkable coulombic interaction occurred between H4K20me2 Arg17 and Glu929. Like hydrogen bond interactions, coulombic interactions between these residues were formed in the last 10 ns of the simulation.

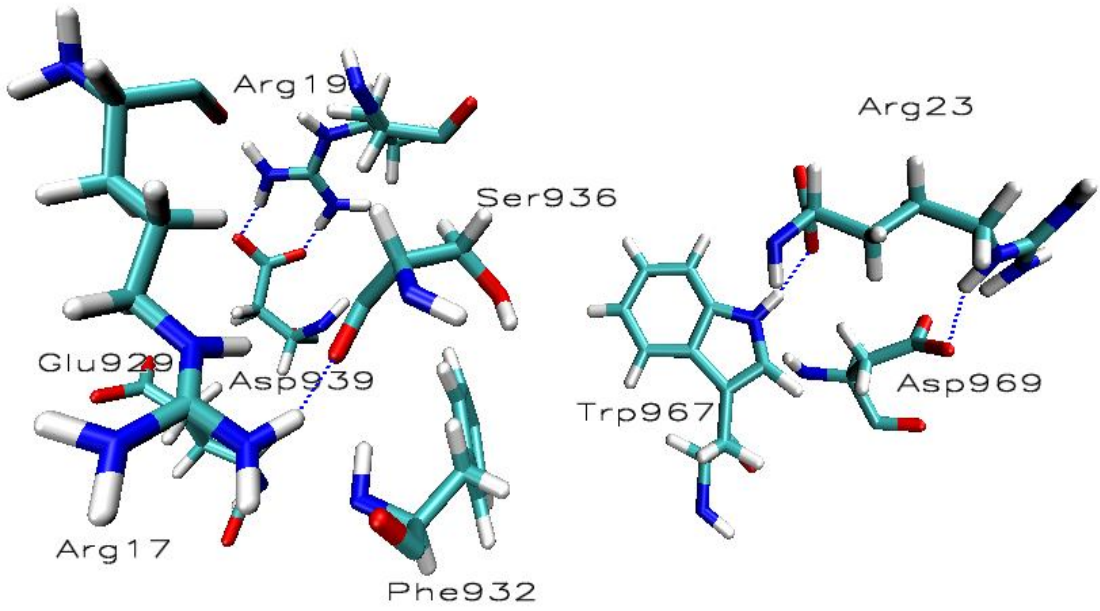


Figure 5.8: Critical residues involved in forming hydrogen bonds and salt bridges between JMJD2A-tudor and H4K20me2 peptide. Hydrogen bonds are shown in dashed blue lines.

5.3.4 JMJD2A Tudor with H3K9me3

The last structure, JMJD2A-tudor binding to H3K9me3, appeared to have the least interactions between the ligand and the receptor. Although the initial structure was obtained from docking and may not represent the exact conformation about the structure, molecular dynamics simulation of the structure gives a suggestion.

In this structure H3K9me3 Arg8 plays a crucial role in binding. Hydrogen atoms on NH1 and NH2 atoms of Arg8 hydrogen bond with Asn940 and Thr968. Interaction with Asn940 suggests that H3K9me3 Arg8 may have a similar binding fashion with H3K4me3 Thr3. Corroborating this idea, Arg8 also formed another strong hydrogen bond with Ser938 like the hydrogen bond between Ser938 and H3K4me3 Thr3 (Figure 5.9). However, the bond between H3K9me3 Arg8 and Ser938 had a higher occupancy of 66.57% and persistent throughout the trajectory. Ala7 on the histone tail was another residue that was found to be interacting with the tudor domain. H1, H2 and H3 atoms on the backbone nitrogen atom formed hydrogen bonds with Asp1001. Since Asp1001 is in the HTD1 lobe of the Tudor domain, this interaction may lean to the unrestrained terminal residues that move freely. Ser10 and Thr11 of histone tail formed another hydrogen bond between each other that prevented the ligand interacting with the receptor via these residues.

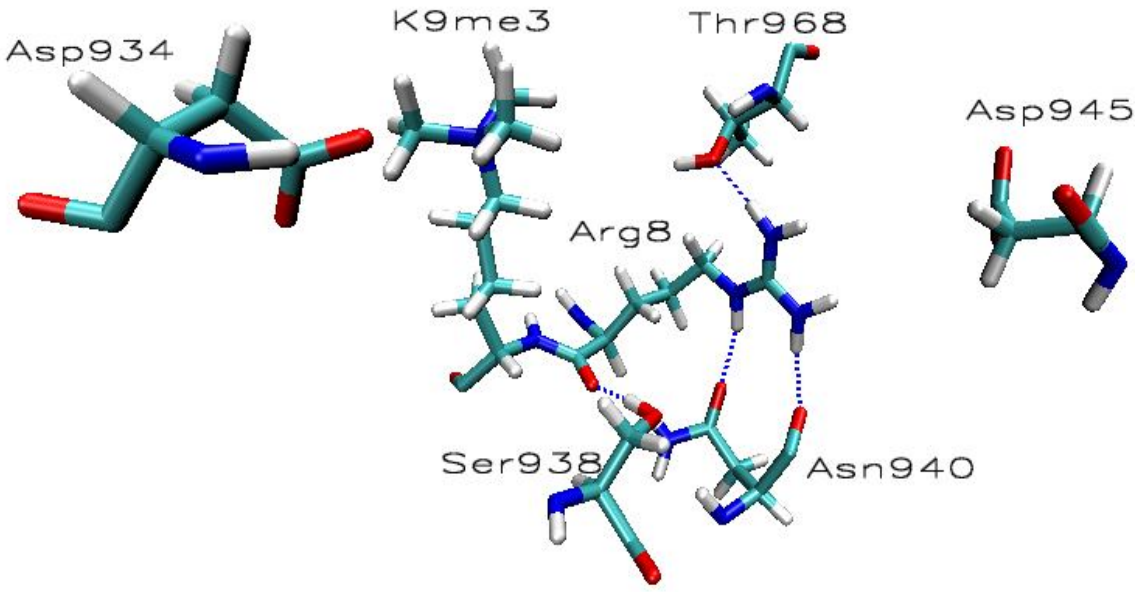


Figure 5.9: Critical residues involved in forming hydrogen bonds and salt bridges between JMJD2A-tudor and H3K9me3 peptide. Hydrogen bonds are shown in dashed blue lines.

The most notable coulombic interaction was formed between trimethylated Lys9 and Asp934. Although Asp945 was formed hydrogen bond and coulombic interactions with Arg966 like the previous two structures, it also formed some salt bridges to Arg8 of histone tail during the first half of the simulations.

Chapter 6

FREE ENERGY ANALYSIS

6.1 Binding Free Energy Decomposition

6.1.1 PB/GB Calculations

Mean energy values for each structure were computed for different dielectric constants with corresponding energy contributions as explained before. Table 6.1 shows the components of the enthalpic contribution to the binding free energy. The non polar contributions and the internal energy contributions, which come from the sum of bond, angle and dihedral energies, constitute a small part of the enthalpy. As expected, the electrostatic and the VDW arise from the MM part and the polar contribution that comes from the PB/GB calculations constitute the major part of the enthalpy.

Since the dielectric constant does not affect the non polar energies in molecular mechanics, the VDW and the internal energies are stable for each structure. As the dielectric constant for the protein increases, the electrostatic energy of the structures and the polar contributions, those come from the PB/GB method, decreases. However, the difference between the electrostatic energy contribution calculated by the MM and the polar contribution calculated by the PB/GB solver gets smaller as the dielectric constant increases. This increase yields a decrease in the enthalpic contribution to the binding free energy.

Table 6.1*: Free energy contributions coming from molecular mechanics, GB and PB calculations

Structure	ε	ΔE^{ele}	ΔE^{vdw}	ΔE^{int}	ΔE^{gas}	ΔG^{SA}(GB)	ΔG^{GB}	ΔG^{MMGBSA}	ΔG^{SA}(PB)	ΔG^{PB}	ΔG^{MMPBSA}
Tudor-H3K4me3	**1**	-433.46	-39.65	4.2	-468.91	-4.16	454.72	**-18.35**	-5.66	458.66	**-15.91**
	2	-216.73	-39.65	4.2	-252.18	-4.16	224.48	**-31.86**	-5.66	226.18	**-31.66**
	4	-108.36	-39.65	4.2	-143.81	-4.16	109.36	**-38.61**	-5.66	109.95	**-38.82**
Tudor-H4K20me3	**1**	-827.38	-45.81	-6.07	-879.26	-4.62	844.3	**-39.58**	-6.44	910.37	**24.68**
	2	-413.69	-45.81	-6.07	-465.57	-4.62	416.81	**-53.38**	-6.44	449.47	**-22.53**
	4	-206.84	-45.81	-6.07	-258.72	-4.62	203.06	**-60.28**	-6.44	219	**-46.16**
Tudor-H4K20me2	**1**	-919.05	-44.44	-4.21	-967.71	-4.64	927.01	**-45.35**	-6.31	951.42	**-22.6**
	2	-459.53	-44.44	-4.21	-508.18	-4.64	457.64	**-55.19**	-6.31	469.63	**-44.86**
	4	-229.76	-44.44	-4.21	-278.42	-4.64	222.95	**-60.11**	-6.31	228.72	**-56**
Tudor-H3K9me3	**1**	-443.3	-40.31	9.63	-473.99	-4.64	461.19	**-17.45**	-6.04	461.96	**-18.07**
	2	-221.65	-40.31	9.63	-252.34	-4.64	227.67	**-29.31**	-6.04	228.2	**-30.18**
	4	-110.83	-40.31	9.63	-141.51	-4.64	110.92	**-35.24**	-6.04	111.28	**-36.27**

* All values in the table are in kcal/mol units.

In all of the four complexes intermolecular coulombic forces and van der Waals interactions favour the ligand binding. Internal energies also favour binding of H4K20me2/3 ligands, whereas disfavour binding of H3K4me3 and H3K9me3 ligands proposing that the conformational changes upon binding lead to internal strains in JMJD2A-tudor- H3K4me3/H3K9me3 complexes [63][64]. Since internal energy consists of bond angle and dihedral terms, one might think of a correlation between RMSD and internal energy. In this study; however, RMSD values are shown to be uncorrelated with resulting internal energies, suggesting that higher RMSD values may not always indicate dispersed internal strains in protein structures. To investigate further, residue averaged RMSF values were seek for any correlation with internal energies and indeed, a positive correlation was observed between them. This finding; hence, shows that higher RMSF values are indicators of bigger internal strains.

The non polar solvation free energy values for the GB model, which was obtained via changing SASA upon binding, contributes favourably to the total binding free energy in all of the complexes. The polar contributions to the solvation free energy for the GB model, on the other hand, considerably disfavours the binding for all complexes and in all dielectric constants. The total electrostatic energies ($\Delta E^{ele}+\Delta G^{GB}$) are positive in the tudor-H3K4me3/H3K9me3 complexes, indicating that overall coulombic forces disfavour binding. In the tudor-H4K20me2/3 complexes the total electrostatic energies decrease from positive to negative values as the dielectric constant increases. The total electrostatic energy is negative when the dielectric constant is 4 in the tudor-H4K20me3 complex, whereas in the tudor-H4K20me2 complex the total electrostatic energies are negative when the dielectric constant is 2 and 4, proposing that the total coulombic interactions slightly favour binding. The compensation of the electrostatic energies with the polar solvation free energy leans to the high cost of desolvation of the uncounterbalanced polar and charged groups upon complex formation. Overall, this proposes that, for all complexes binding is

mainly driven by favourable van der Waals interactions. The non polar contributions to the total solvation free energy and the molecular mechanical internal energies have a less significant contribution to the binding.

The non polar contributions to the solvation free energy values for the PB model, favour binding as in the GB model. As expected the polar contributions to the solvation free energy that is obtained by PB model are unfavourable to the ligand binding for all complexes. The total electrostatic interaction energies are positive for all complexes and in all dielectric constants except the tudor-H4K20me2 complex under the dielectric constant 4 which has a quite small contribution of -1.04 kcal/mol. Hence, it is observed that binding to ligands are mostly electrostatically unfavourable due to the desolvation penalty upon ligand binding. PB method also supports that favourable binding is mainly driven by van der Waals interactions between the receptor end the ligand. Furthermore, the non polar contributions to the solvation free energy and the internal energies also contributes to the binding slightly. Interestingly, the tudor-H4K20me3 complex under the dielectric constant 1 appears to be unfavourable in binding to the ligand. Since experimental studies show that this binding is energetically favourable, the calculated result seems to be strange. An explanation to this positive value of 24.68 kcal/mol is that the number of charged residues and the highly polar groups in H4K20me3 peptide might have leant to high desolvation penalties in PB calculations so that even van der Waals values could not have compensated the unfavourable electrostatic contributions.

The enthalpies that were obtained via GB continuum solvent model are in a good agreement with the ones that are obtained via PB continuum solvent model for the tudor-H3K4me3 and the tudor-H4K9me3 complexes. However, for the tudor-H4K20me2/3 complexes the enthalpies that were obtained via PB and GB methods slightly differ from each other. To figure out the underlying reason for that, differences between the parameter sets for PB and GB methods should carefully be investigated. Because GB method is

known to be less accurate in estimating the polar contribution to the solvation energy upon binding, PB calculations is expected to give a better representation for the systems. Therefore, one might suppose that this disagreement between PB and GB methods depends on the inaccuracy of the GB method. However, the good agreement between the methods for the other two complexes suggests that, the problem leans to the highly polar residues in the H4K20 ligand. It has been previously observed that PB calculations with PARSE radii have higher desolvation penalties compared to the calculations with modified Bondi radii. PARSE radii values are smaller than modified Bondi radii values; hence they bear strong polarization effect for the polar groups. Since the polar groups on the ligand is buried inside the complex upon complex formation, strong polarization leads to high desolvation penalties when PARSE radii is applied. Furthermore, it was previously proposed that the GB model favours coulombic interactions between charged groups leaning to overestimation of binding affinities more than the PB model does [65].

6.1.2 Entropy Calculations

The continuum solvent models estimate the free energy comprising the contribution of the solvent entropies; thus changes in the solute entropy should be excluded from the final enthalpies. The entropic contributions emanates from the conformational changes in rotational, translational and vibrational degrees of freedom of solvent upon complex formation (Table 6.2). The loss in translational and rotational degrees of freedom was calculated based on classical statistical mechanics; whereas, the loss in vibrational degrees of freedom was calculated using normal mode analysis. Standard errors of the entropic contributions entirely arose from the vibrational degrees of freedom by around 1kcal/mol which is highly reasonable in terms of internal accuracy of the snapshots. It has been previously proposed by Stoica et al. [66] and Chen et al. [67] that, enthalpic and entropic

contributions are anti correlated. Opposing this finding, further investigation of the entropic values together with enthalpic contributions suggested that, enthalpy and entropy contributions are not anticorrelated for all systems. In this study; although, there may appear a possible correlation between enthalpic and entropic contributions for the tudor-H3K4me3 and the tudor-H4K20me2/3 complexes, the tudor-H3K9me3 complex abolishes this correlation. Strikingly, we found that instead of enthalpy, electrostatic and var der Waals contributions were anticorrelated with entropy contributions in our systems. However, omission of configurational degrees of freedom that arose from side chain reorganisation, may lead significant changes in the final entropic contribution [68]. Therefore, one can overlook a possible correlation of entropic contribution with enthalpic contribution without the addition of configurational degrees of freedom. Moreover, a variation in the entropic contribution may lead to crucial outcomes in the binding free energy.

Table 6.2*: Contributions of the entropic components to the binding free energies.

Structure	$-T\Delta S_{trans}$	$-T\Delta S_{rot}$	$-T\Delta S_{vib}$	$-T\Delta S_{tot}$
Tudor-H3K4me3	-13.73	-12.06	-0.28	-26.08
Tudor-H4K20me3	-13.91	-12.55	-5.97	-32.43
Tudor-H4K20me2	-13.9	-12.63	-9.76	-36.29
Tudor-H3K9me3	-13.62	-11.93	-3.72	-29.27

* All values are in kcal/mol units.

6.2 Binding Free Energy Analysis

Combined with entropic contributions, enthalpic contributions bear the binding free energy. Upon the use of PB or GB model for obtaining the polar solvation contribution, not surprisingly resulting binding free energies differ from each other. Since the resulting binding free energies are positive; hence indicating unfavourable interactions, the enthalpic contributions under the dielectric constant 1 and 2 are not going to be discussed any further. Because there is not any experimental data for two of the four complexes, comparison with experimental data involves only two of the interactions. In this respect, discussions in model comparisons will be based on the available data.

For the tudor-H3K9me3 and the tudor-H3K4me3 complexes ΔG_{PB} and ΔG_{GB} values are in a good agreement, whereas for the tudor-H4K20me2/3 complexes the values are considerably different. Despite the tudor-H4K20me2 complex is the most favourable with distinction by a binding free energy of -19.71 kcal/mol from ΔG_{PB} values, from ΔG_{GB} values the tudor-H4K20me3 complex appears to be the most favourable by a binding free energy of -27.85 kcal/mol. Owing to the neglected contributions to entropy and enthalpy, the dielectric constant, the bond radii and the model choice for solving the solvation free energy, calculated binding free energies deviate from the experimental values. Compared with the experimental binding free energies, it is seen from table 6.3 that, ΔG_{PB} values underestimate the experimental binding free energy by 5 kcal/mol and ΔG_{GB} values underestimate the experimental binding free energy by 4-19 kcal/mol. In this study; hence, ΔG_{PB} values seem to be more realistic and more consistent with the experimental values. However, it is important to emphasize that this finding may not always be true for every system.

Table 6.3*: Binding free energy components of the structures calculated from GB and PB methods and experimental disassociation constants.

Structure	ΔG^{MMGBSA}	ΔG^{MMPBSA}	$-T\Delta S_{tot}$	ΔG_{GB}	ΔG_{PB}	ΔG_{exp}
Tudor-H3K4me3	-38.61	-39.52	-26.08	-12.53	-13.44	-8.64
Tudor-H4K20me3	-60.28	-46.16	-32.43	-27.85	-13.73	-8.77
Tudor-H4K20me2	-60.11	-56	-36.29	-23.82	-19.71	n/a
Tudor-H3K9me3	-35.24	-36.27	-29.27	-5.97	-7	n/a

* All values are in kcal/mol units.

6.3 Analysis of Hot Spots

In this study, the enthalpic contribution to the binding free energy was decomposed into its residual components and the residual components were decomposed into pair-wise components. From the contribution of each of the residues to the binding free energy, residues that have significance in binding were identified. Since the dielectric constant 4 is chosen for its best representation of binding free energy, only the decomposition results for the dielectric constant 4 is presented. The residues in the ligand and in HTD-2 of the receptor which have a contribution of less than -1.0 (kcal/mol) to the enthalpic contribution the total binding free energy were defined as hot spots. Since decomposition of the entropic contribution on per-residue basis to the binding free energy was not available, hot spots were determined on the basis of the enthalpy terms. It should be noted that, inclusion of entropic contributions might give much more accurate results.

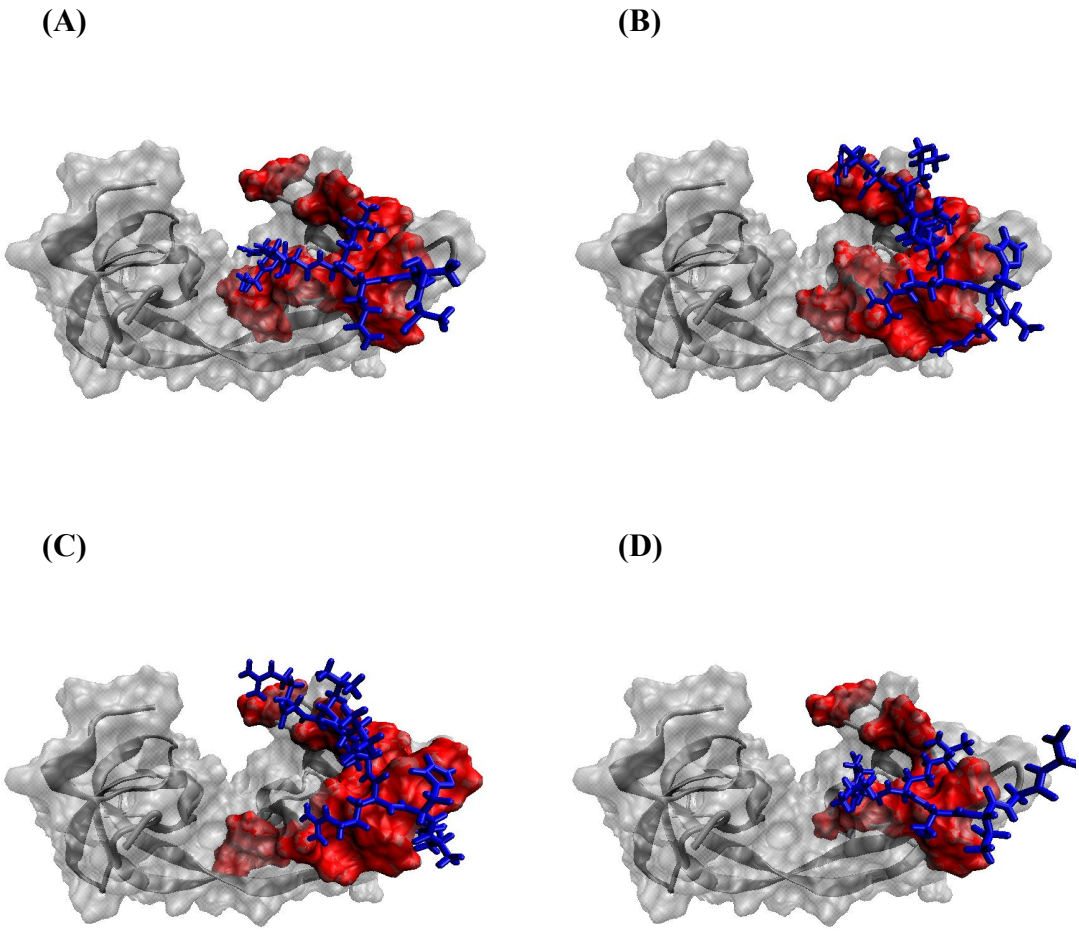


Figure 6.1: Molecular surface representation of JMJD2A-tudor. The hot spot residues in the receptor are shown with red colour. Shown in licorice representation with blue colour, the ligand residues of H3K4me3 (A), H4K20me3 (B), H4K20me2 (C) and H3K9me3 (D) structures are represented in the figure.

6.2.1 JMJD2A Tudor with H3K4me3

As seen in table 6.4, upon binding of JMJD2A-tudor to the H3K4me3 histone tail, Asn940 was found to be having the highest binding free energy contribution of -3.88 kcal/mol. The high contribution to the overall favourability was mainly driven by van der Waals and electrostatic interactions between Asn940 and Ala1, Arg2 and Thr3 of the histone ligand with -2.26, -1.07 and -1.75 kcal/mol ΔG contributions respectively. Indeed, Asn940 was found to be one of the most critical residues in binding to H3K4me3 ligand experimentally.

Neighbouring the trimethyllysine residue, Phe932, Trp967 and Tyr973 contributions were found to be remarkable. As shown in table 6.4, the total ΔG contributions to the overall binding free energy are -3.09, -3.61 and -1.06 kcal/mol respectively. The three of the residues formed strong van der Waals interactions with trimethyllysine residue via their aromatic side chains. Asp969, another significant residue, also contacted with Trp967 upon complexation, hence favouring the binding. Despite Tyr973 shows a relatively small contribution to the total enthalpy, contribution of the trimethyllysine in free energy was -2.26 kcal/mol which is quite significant. Due to the unfavourable van der Waals interactions with Trp967, resulting contribution of Tyr973 appears to be smaller.

Having contributions less than -2.0 kcal/mol, Ser936, Phe937 and Leu941 were observed to be very remarkable. Ser936 made van der Waals contacts with trimethyllysine, Gln5 and Thr6 in methylated ligand. Phe937 also favours binding, considering the high desolvation penalty coming from the polar contribution to the solvation free energy. It formed favourable side chain and backbone van der Waals contacts and favourable backbone electrostatic interactions with trimethyllysine and Gln5 of the ligand. Leu941 contacted with Arg2 of the methylated ligand through side chain and backbone van der Waals interactions. Interestingly contribution of internal energies to the free energy, by a

value of -0.95 kcal/mol, is very significant for Leu941. This proposes that, conformational changes lead to internal relaxation of Leu941 upon complex formation.

Phe927 and Tyr942 together formed side chain van der Waals interactions with Ala1 of the histone tail. Moreover, these two residues slightly favoured binding by also interacting with each other. Another worthy contribution came from Ser938 which contacted with the trimethyllysine residue through side chain and backbone van der Waals interactions and Thr3 of the methylated lysine residue via electrostatic interactions, thus contributing by -1.11 and -0.91 kcal/mol respectively.

Table 6.4*: Free energy decomposition of hot spot residues in JMJD2A-tudor-H3K4me3.

Molecule	Residue	INT	VDW	ELE	GB	GBSUR	GBTOT
JMJD2A-tudor	PHE927	-0.26	-1.64	0.43	-0.31	-0.13	-1.92
	PHE932	0.04	-2.96	0.12	-0.20	-0.09	-3.09
	SER936	-0.08	-2.13	-0.49	0.63	-0.25	-2.31
	PHE937	-0.08	-2.70	-1.33	1.14	-0.32	-3.29
	SER938	0.41	-1.76	-0.31	0.13	-0.14	-1.66
	ASN940	0.06	-2.87	-4.38	3.67	-0.37	-3.88
	LEU941	-0.95	-1.49	-0.15	-0.05	-0.06	-2.69
	TYR942	0.26	-1.41	0.18	-0.00	-0.19	-1.16
	ASP945	-0.72	-0.31	-11.87	11.64	-0.08	-1.33
	TRP967	-0.25	-3.18	-0.55	0.78	-0.41	-3.61
	ASP969	-1.07	-0.78	-9.90	10.05	-0.22	-1.92
	TYR973	-0.18	-0.64	-0.50	0.36	-0.10	-1.06
H3K4me3	ALA1	0.23	-2.27	-26.93	27.53	-0.48	-1.92
	ARG2	4.09	-4.09	-35.08	32.43	-0.63	-3.28
	THR3	-1.45	-1.15	-1.42	1.33	-0.26	-2.94
	M3L4	-0.45	-7.42	-29.75	28.38	-0.72	-9.95
	GLN5	-0.54	-1.84	-2.63	2.71	-0.35	-2.65

* All values are in kcal/mol units.

Strikingly, in this study, the difference in free energy of Asn945 upon complex formation, showed up with a relatively low value (-1.33 kcal/mol), although the residue was known to be very crucial in binding to the ligand. However, dominated by side chain electrostatic interactions, the notable contribution to the free energy difference of Arg2 in the histone tail, which was a value of -3.26 kcal/mol, suggested that Asn945 had a major role in binding, in spite of the fact that the high desolvation penalty cancelled the overall electrostatic contribution term.

The first five residues in the ligand are shown to be highly important in complex formation with very big values of free energy differences. Without counting the trimethyllysine, among all the residues in the ligand Arg2 is the most crucial one. It should be noted that, Arg2 formed many interactions with the receptor molecule, thus favouring the binding.

6.2.2 JMJD2A Tudor with H4K20me3

In experimental studies, trimethyllysine residue in tudor-H4K20me3 complex was observed to be caged by the residues Phe932, Trp967 and Tyr973. Shown in table 6.5, favouring the complexation, the binding free energy differences of the residues are very low. Aromatic side chains of these amino acids formed a hydrophobic pocket for the binding of the methylated lysine in the ligand. Side chains of Phe932 and the trimethyllysine residue formed favourable van der Waals interactions and together with the negative contribution of the polar desolvation energy, the trimethyllysine contributed to the free energy by -3.14 kcal/mol upon binding. Trp967 also favoured binding via electrostatic, van der Waals and hydrogen bond interactions that were established throughout the simulation. The total contribution of Trp967 to the enthalpy, which is -4.93 kcal/mol, came from the van der Waals interactions that were formed with the trimethyllysine. As in the

tudor-H3K4me3 complex, the energetic contribution of Tyr973 to the total enthalpy appears to be small, compared to the previous two residues. However, the contribution of the trimethyllysine by a value of -2.51 kcal/mol proposes that, Tyr973 has a remarkable role in binding.

Ser936, Phe937 and Asp 969 were found to be significant, in terms of their negative energetic contributions to the enthalpy upon complex formation. With a value of -2.32 kcal/mol in contribution, Ser936 favoured binding by van der Waals contacts with His18 of the ligand and the trimethyllysine residue. Intriguingly, Phe937 displayed a very low energetic value of -4.76 kcal/mol to the total enthalpic contribution of the binding free energy. Mostly dominated by van der Waals contacts, analyzing the energetic contributions suggest that many interactions occurred between the Arg17, His18, Arg19 and the trimethyllysine residue of the histone ligand. Considerably big contribution values coming from the Arginine residues of the H4K20me3 ligand (-3.59 and -3.55 kcal/mol for Arg17 and Arg19 respectively) signifies that, Phe937 is a crucial residue in binding to the ligand. Asp969, with a fairly important total contribution, formed electrostatic contacts with Gln971 and van der Waals contacts with Trp967 upon complex formation. The relatively big value of internal contribution suggests that internal strain was removed upon complex formation. Moreover, salt bridges that were formed with Arg23 of the histone peptide, is demonstrated with a contribution of -1.25 kcal/mol.

Shown in table 6.5, Glu929 contributes to the enthalpic contribution of the binding free energy by a value of -1.69 kcal/mol which is dominated by the electrostatic contribution coming mostly from the hydrogen bond and coulombic interactions with Arg17 in the ligand. It should be noted that this residue might be significant in the binding, since a very low contribution comes from Arg17 by a value of -2.38 kcal/mol.

Table 6.5*: Free energy decomposition of hot spot residues in JMJD2A-tudor-H4K20me3.

Molecule	Residue	INT	VDW	ELE	GB	GBSUR	GBTOT
JMJD2A-tudor	PHE927	0,17	-1,37	0,09	0,05	-0,14	-1,21
	GLU929	-0,30	0,11	-20,98	19,58	-0,11	-1,69
	PHE932	-0,05	-2,25	-0,38	0,24	-0,09	-2,53
	SER936	-0,19	-1,86	-0,58	0,57	-0,25	-2,32
	PHE937	0,95	-5,05	-1,26	1,13	-0,52	-4,76
	SER938	-0,14	-1,65	0,48	-0,17	-0,11	-1,58
	ASP939	0,42	-0,34	-20,92	19,70	-0,21	-1,35
	LEU941	-0,18	-0,86	-0,50	0,26	-0,02	-1,29
	ASP945	-0,66	-0,35	-9,73	9,68	-0,01	-1,07
	GLU961	-0,03	-0,36	-8,39	7,90	-0,05	-0,93
	TRP967	-0,75	-3,54	-1,56	1,40	-0,49	-4,93
	THR968	0,31	-0,80	-1,10	0,52	0,06	-1,01
	ASP969	-0,99	-1,72	-13,77	14,01	-0,39	-2,86
	GLY970	-0,59	-0,42	-0,41	0,31	-0,12	-1,23
	TYR973	-0,48	-0,63	-0,28	0,17	-0,09	-1,30
H3K4me3	ARG17	-0,14	-2,23	-55,33	52,94	-0,35	-5,12
	HIS18	-0,06	-1,50	-0,15	0,21	-0,07	-1,57
	ARG19	-0,70	-4,18	-32,19	30,46	-0,57	-7,17
	M3L20	0,42	-8,14	-31,31	30,39	-0,83	-9,47
	LEU22	-1,23	-0,47	-0,34	0,51	-0,09	-1,63
	ARG23	-1,91	-1,45	4,16	-2,94	-0,17	-2,31

* All values are in kcal/mol units

Despite Asp939 was known to be crucial in complex formation, a value of -1.35 kcal/mol free energy difference upon binding was observed in this study. Further investigation of the components of the contribution suggests that the residue has a remarkable role in binding. Asp939 favoured binding of Arg17 and Arg19 by -0.53 and -2.57 kcal/mol respectively and the arginine residues also contributed to the free energy difference of Asp939 by -0.53 and -2.52 kcal/mol. Favourability of Asp939 was driven by electrostatic interactions which were comprised of the hydrogen bonds and salt bridges,

even though a big portion of the electrostatic term was cancelled by the desolvation penalty.

Located at nearby of the binding area of the receptor Ser938, Leu941, Thr968 and Gly970 were mainly occupied in making hydrophobic interactions with the residues Arg966 and Trp967; hence favouring the binding . Among these residues, Ser938 formed weak side chain and van der Waals interactions with Arg19 and the methylated lysine residue in the tail peptide. Thr968, another remarkable residue, also formed weak backbone van der Waals contacts and electrostatic interactions with Arg23 residue in the ligand.

Noteworthy to mention that although Asp945 is defined as a hot spot in this study, the residue was not observed to have any contribution coming from the peptide ligand. The contributions were mainly supplied through side chain van der Waals forces with the receptor residues. Having less significance, contributions to the Phe927 and Glu961 arose from the van der Waals forces with the receptor residues Asn940, Leu941 and Tyr942, and Pro957 and Val963 respectively.

6.2.3 JMJD2A Tudor with H4K20me2

The side chain of dimethyllysine residue in the tudor-H4K20me2 complex was caged by aromatic side chains of Phe932, Trp967 and Tyr973 like the trimethyllysine residue in the tudor H4K20me3 complex. Having bigger contributions coming from the van der Waals forces upon complexation, the free energy differences for these amino acids lean to the hydrophobic interactions with the dimethyllysine. Considerably, very low value of Trp967 free energy difference arose not only from the interactions with the dimethyllysine, but also from the hydrogen bond interactions with the Arg23 of the ligand. It should be remembered that, for the trimethyllysine residue, Thr968 was occupied in interactions with Arg23 residue. The remaining residues were seen to be having the same interactions in both

of the cases. Overall, resulting values suggest that binding to the dimethyllysine is much favoured by the aromatic pocket residues, compared to the trimethyllysine amino acid in H4K20me3 structure (Table 6.6).

Table 6.6*: Free energy decomposition of hot spot residues in JMJD2A-tudor-H4K20me2.

Molecule	Residue	INT	VDW	ELE	GB	GBSUR	GBTOT
JMJD2A-tudor	PHE927	-0.22	-1.46	0.35	-0.17	-0.11	-1.61
	GLU929	-0.07	0.05	-14.81	14.00	-0.03	-0.86
	PHE932	0.09	-2.34	-0.83	0.59	-0.08	-2.56
	ASP934	0.25	-0.50	-18.93	18.28	-0.09	-0.98
	GLY935	-0.23	-1.09	-0.62	0.76	-0.14	-1.31
	SER936	-0.12	-2.76	-1.09	0.97	-0.27	-3.27
	PHE937	0.32	-4.16	-1.26	1.18	-0.44	-4.36
	SER938	-0.09	-1.30	0.30	-0.08	-0.07	-1.24
	ASP939	0.18	-0.08	-17.07	16.01	-0.08	-1.04
	GLU961	0.01	-0.41	-8.72	8.13	-0.07	-1.06
	TRP967	-1.31	-2.77	-2.52	1.95	-0.35	-5.00
	ASP969	-1.04	-1.57	-14.63	14.71	-0.40	-2.93
	TYR973	-0.53	-0.76	-0.41	0.29	-0.13	-1.54
H3K4me3	ARG17	0.01	-3.76	-56.14	54.08	-0.52	-6.33
	HIS18	-0.21	-3.08	-0.13	0.43	-0.28	-3.27
	ARG19	-0.55	-2.70	-29.68	28.38	-0.25	-4.80
	M2L20	0.04	-9.26	-32.25	31.26	-1.04	-11.25
	LEU22	-0.69	-1.15	-0.90	1.02	-0.12	-1.84
	ARG23	-0.40	-2.61	0.97	-0.25	-0.38	-2.67

* All values are in kcal/mol units

With -2.56, -3.27 and -4.36 kcal/mol energetic values respectively, Ser936, Phe937 and Asp969 contributed to the total enthalpic change significantly in binding to H4K20me2 histone peptide. In addition to the pair wise contributions coming from His18 and the methylated lysine residue as in the trimethylated structure, Arg17 also contributed to Ser936 in the tudor-H4K20me2 structure. As previously proposed in the tudor-H4K20me3

structure, Phe937 has also a very remarkable role in binding to H4K20me2 ligand, revealed by the contributions that were mostly dominated by arginine residues and coming from the Arg17, His18, Arg19 and the dimethyllysine.

After complex formation, favourable energy values leaning to interactions within the receptor residues were observed in the tudor-H4K20me2 complex. As previously announced for the tudor liganded to trimethylated H4K20 structure, Phe927 and Ser938 formed favourable van der Waals contacts with the receptor amino acids in the present structure. In addition to this, another important residue, Glu961 made strong coulombic interactions with the receptor residues as well.

In this study Glu929, Asp934, Gly935 and Asp939 are introduced as hot spots for the complexation of the tudor with H4K20me2 peptide, although first two of them display values higher than -1 kcal/mol. However, considering their pair-wise contributions, they are thought to have importance in binding. Coming from van der Waals and electrostatic interactions Arg17 of the ligand contributed to the free energy difference of Glu929 and Gly935 (-1.02 and -1.11 kcal/mol respectively); hence emphasizing the role on binding to the receptor. Gly935 was also occupied in van der Waals interactions with His18. Asp934 comes into prominence with its favourable interactions with the dimethyllysine which were mainly dominated by hydrogen bonds. Asp934 interacted with the trimethyllysine in the H4K20 ligand without forming hydrogen bonds; therefore the resulting energetic value is lower for the dimethylated structure. In terms of binding to the ligand another crucial residue, Asp939 was occupied in interacting with arginine residues in the tail peptide. Arg19 contributed to the favourability of the residue by a value of -1.68 kcal/mol, whereas Arg17 contributed by a value of -0.46 kcal/mol through hydrogen bonds and coulombic interactions.

6.2.4 JMJD2A Tudor with H3K9me3

The final structure that was investigated in this study is the tudor-H3K9me3 complex. As appeared in the former structures, the trimethyllysine residue in the H3K9me3 peptide was caged by the aromatic side chains of the receptor residues Phe932, Trp967 and Tyr973. Despite the first two of the residues are defined as hot spots, the former residue is not introduced as a hot spot. The relatively high value of Tyr973, which is above -1 kcal/mol, arose from the unfavourable interactions with Asp933, Asp934 and Trp967. It should be noted that Tyr973 was crucially involved in binding of the trimethyllysine residue to the receptor, although it does not exist in the list (see Table 6.7). With large negative values, Phe932 and Trp967 contributed to the binding favourably. Phe932 contacted with the trimethyllysine residue via van der Waals and electrostatic forces, whereas Trp967 made dominant van der Waals contacts with the trimethyllysine residue, as well as with the Arg8 in the H3K9me3 ligand.

Ser936 and Phe937 were found to be highly remarkable, considering their contributions upon complexation. Ser936 contacted with the trimethyllysine, Ser10, Thr11 and Gly12 through favourable van der Waals forces. Having favourable interactions with many of the residues in the ligand, Ser936 appears to be vital in complexation. Backbone and side chain of Phe937, were also involved in favourable van der Waals interactions with the methylated lysine in H3K9me3 peptide, Ser10 and Thr11, indicating the importance of the residue in binding to the ligand.

Table 6.7*: Free energy decomposition of hot spot residues in JMJD2A-tudor-H3K9me3.

Molecule	Residue	INT	VDW	ELE	GB	GBSUR	GBTOT
JMJD2A-tudor	PHE932	-0.01	-2.79	-0.02	-0.05	-0.09	-2.96
	SER936	-0.34	-2.32	-1.04	0.98	-0.27	-2.99
	PHE937	-0.04	-2.61	-1.40	1.18	-0.31	-3.18
	SER938	-0.03	-1.52	0.14	0.10	-0.13	-1.44
	ASN940	0.14	-2.35	-2.53	2.38	-0.36	-2.72
	LEU941	-0.74	-0.90	-0.04	-0.15	-0.05	-1.88
	TRP967	0.15	-3.54	-0.39	0.44	-0.46	-3.80
	ASP969	-1.01	-0.61	-6.61	6.46	-0.12	-1.88
	GLY970	-0.50	-0.47	-0.35	0.25	-0.14	-1.21
	HIS981	-1.14	-0.84	-0.58	0.73	-0.13	-1.95
H3K4me3	ARG8	-0.06	-4.13	-27.56	26.66	-0.75	-5.84
	M3L9	-0.19	-8.06	-32.30	30.84	-0.97	-10.68
	SER10	-0.59	-0.49	-1.23	1.55	-0.27	-1.04

* All values are in kcal/mol units

Shown in table 6.7, contribution of Asn940 is relatively very low by a value of -2.72 kcal/mol. It should be recalled that Asn940 was also occupied in favourable interactions and introduced as a hot spot in the tudor-H3K4me3 complexation. In the tudor-H3K9me3 complex, Asn940 made van der Waals and hydrogen bond contacts with Phe927 and the ligand residue Arg8 by contribution values of -0.69 and -2.24 kcal/mol respectively. Recalling that Asn940 interacted with Thr3 of H3K4me3 ligand, for the present structure Arg8 takes the place of Thr3.

Defined as a hot spot in the former three structures, Ser938 was also involved in favourable interactions with the tudor-H3K9me3 complex. The receptor residues that contacted with Ser938 upon liganding to H3K9me3 peptide suggest that the role of this residue in binding is very similar to that of liganding to H3K4me3 peptide. In the present structure, Ser938 formed strong hydrogen bond interactions with Arg8, favourable van der

Waals interactions with the trimethyllysine9, and some weak interactions with Ser10, thus favouring the complexation.

Along with Ser938, residues Leu941, Asp969 and Gly970 are beheld to be worthy in complex formation with the favourable contributions coming from the Trp967. In addition to the contribution of Trp967, Leu941 was also favoured by the van der Waals interactions with Arg8 in the ligand. Asp969, as seen in the previous complexes, formed strong interactions with Gln971 by a contribution value of -0.95 kcal/mol, depicting the favourability of the binding. Internal energy difference upon binding suggests that internal strains of Asp969 were removed after complex formation. In this respect, defined as a hot spot His981 is another instance for representing a low negative energetic value arose from the internal relaxation. Gly970 was involved in favourable interactions with Arg966 and Gln971 along with Trp967 and had not any direct involvement in ligand interactions.

Chapter 7

CONCLUSIONS

In this study, we performed 25ns fully unrestrained molecular dynamics simulations of the tudor domains of JMJD2A complexed with H3K4me3, H4K20me3, H4K20me2 and H3K9me3 histone tails along with the free structures. For the non standard trimethyllysine and dimethyllysine residues, parameters compatible to the Duan et al. force field were generated using quantum mechanical techniques. Docking simulations were carried out for JMJD2A-tudor-H3K9me3 complex before the simulations of the structure, since there was no available initial structure determined by the experiments.

Revealed by the MD simulations, different distinct binding modes of the peptides were clarified in detail. We showed that there are two distinct binding modes adopted by the histone ligands that were investigated in this study. The important hydrogen bonds and coulombic interactions between the receptor and the ligand peptides were elucidated and discussed in detail.

Upon investigation of the dynamics of the systems, we found that the tudor domains move towards and against each other periodically. Moreover, suggested by the changes in the radius of gyration values throughout the trajectories, we observed that the behaviour of the global motions change upon binding of the ligands. Further inquiry was performed to analyze the flexibility of the structures by measuring the RMSF values obtained via MD simulations.

From the last 24 ns of the simulations, 2400 snapshots were extracted with equally spaced 10 ps time intervals. Utilizing the snapshots, we calculated the enthalpic contributions to the binding free energies of the four of the complexed structures conducting the three trajectory MM-PBSA/GBSA approaches with varying internal dielectric constants. Convergence analysis of the mean values and the standard errors suggested that the convergence of the GB model is better than the PB model in our systems. In addition, combined with the internal dielectric constant of 4, energy values were observed to be more realistic and showing a better convergence with respect to the other internal dielectric constants.

Binding free energies were obtained after the removal of the entropic terms obtained by the molecular mechanics and normal mode calculations. Entropic and enthalpic contributions to the binding free energies were decomposed into their constituents. We showed that recognition of the peptides by the JMJD2A-tudor was driven by favourable van der Waals interactions, whereas the favourable electrostatic terms were shown to be cancelled by the high desolvation penalties. Decomposition of the entropic determinants revealed that, vibrational degrees of freedom bear the difference in the entropic terms of the structures.

Of the four structures, the one liganded to H4K20me3 appeared to be the most favoured in terms of the relatively low value of the resulting binding free energy. We discovered that the trimethylated H4K20 bound to the tudor domain of JMJD2A structure were more favourable than the dimethylated of the same structure. Nevertheless no significant difference was showed up by means of the enthalpic energetics between the trimethylated and the dimethylated structure; high entropic cost of the dimethylated structure upon binding resulted in a high relative binding free energy compared to the trimethylated structure. Low agreement of the calculated binding free energies with the experimental values was explained by the determinants not included to the enthalpic and entropic

contributions such as conformational entropy. The residues having a contribution less than 1 kcal/mol to the free energy of binding were defined as hotspots and the critical hotspots were discussed in detail. Along with the hotspots unrevealed in available experimental data, consistent with the experiments energetic contributions of the reported significant residues were uncovered.

The present study establishes the missing gaps in the recognition of the H3K4me3 and H4K20me3 peptides and unravels the recognition of the H3K9me3 and the H4K20me2 peptides. While most of the information in this thesis clarifies many points in the recognition of the peptides, yet selective recognition of the histone peptides by the tudor domains of JMJD2A remains to be established indicating that more study is required.

APPENDIX

A.1 AMBER Library File for Dimethyllysine

```
!!index array str
 "M2L"
!entry.M2L.unit.atoms table  str name  str type  int typex  int resx  int
flags  int seq  int elmnt  dbl chg
 "N" "N" 0 1 131075 1 7 -0.480100
 "H" "H" 0 1 131075 2 1 0.296600
 "CA" "CT" 0 1 131075 3 6 0.012700
 "HA" "H1" 0 1 131075 4 1 0.081800
 "C" "C" 0 1 131075 5 6 0.670200
 "O" "O" 0 1 131075 6 8 -0.591300
 "CB" "CT" 0 1 131075 7 6 -0.009700
 "HB1" "HC" 0 1 131075 8 1 0.026900
 "HB2" "HC" 0 1 131075 9 1 0.026900
 "CG" "CT" 0 1 131075 10 6 -0.005100
 "HG1" "HC" 0 1 131075 11 1 0.013900
 "HG2" "HC" 0 1 131075 12 1 0.013900
 "CD" "CT" 0 1 131075 13 6 0.002900
 "HD1" "HC" 0 1 131075 14 1 0.032300
 "HD2" "HC" 0 1 131075 15 1 0.032300
 "CE" "CT" 0 1 131075 16 6 -0.074700
 "HE1" "HC" 0 1 131075 17 1 0.108700
 "HE2" "HC" 0 1 131075 18 1 0.108700
 "NZ" "NA" 0 1 131075 19 7 0.022300
 "HZ" "H" 0 1 131075 20 1 0.300500
 "CZ1" "CT" 0 1 131075 21 6 -0.184900
 "HZ11" "H1" 0 1 131075 22 1 0.130000
 "HZ12" "H1" 0 1 131075 23 1 0.130000
 "HZ13" "H1" 0 1 131075 24 1 0.130000
 "CZ2" "CT" 0 1 131075 25 6 -0.184900
 "HZ21" "H1" 0 1 131075 26 1 0.130000
 "HZ22" "H1" 0 1 131075 27 1 0.130000
 "HZ23" "H1" 0 1 131075 28 1 0.130000
!entry.M2L.unit.atomspertinfo table  str pname  str ptype  int ptypex
int pelmnt  dbl pchg
 "N" "N" 0 -1 0.0
 "H" "H" 0 -1 0.0
 "CA" "CT" 0 -1 0.0
```

```
 "HA" "H1" 0 -1 0.0
 "C" "C" 0 -1 0.0
 "O" "O" 0 -1 0.0
 "CB" "CT" 0 -1 0.0
 "HB1" "HC" 0 -1 0.0
 "HB2" "HC" 0 -1 0.0
 "CG" "CT" 0 -1 0.0
 "HG1" "HC" 0 -1 0.0
 "HG2" "HC" 0 -1 0.0
 "CD" "CT" 0 -1 0.0
 "HD1" "HC" 0 -1 0.0
 "HD2" "HC" 0 -1 0.0
 "CE" "CT" 0 -1 0.0
 "HE1" "HC" 0 -1 0.0
 "HE2" "HC" 0 -1 0.0
 "NZ" "NA" 0 -1 0.0
 "HZ" "H" 0 -1 0.0
 "CZ1" "CT" 0 -1 0.0
 "HZ11" "H1" 0 -1 0.0
 "HZ12" "H1" 0 -1 0.0
 "HZ13" "H1" 0 -1 0.0
 "CZ2" "CT" 0 -1 0.0
 "HZ21" "H1" 0 -1 0.0
 "HZ22" "H1" 0 -1 0.0
 "HZ23" "H1" 0 -1 0.0
!entry.M2L.unit.boundbox array dbl
 -1.000000
 0.0
 0.0
 0.0
 0.0
!entry.M2L.unit.childsequence single int
 2
!entry.M2L.unit.connect array int
 1
 5
!entry.M2L.unit.connectivity table  int atom1x  int atom2x  int flags
 1 3 1
 1 2 1
 3 7 1
 3 4 1
 3 5 1
 5 6 1
 7 10 1
 7 9 1
 7 8 1
 10 13 1
 10 11 1
```

```
 10 12 1
 13 16 1
 13 14 1
 13 15 1
 16 19 1
 16 17 1
 16 18 1
 19 21 1
 19 25 1
 19 20 1
 21 23 1
 21 22 1
 21 24 1
 25 26 1
 25 27 1
 25 28 1
!entry.M2L.unit.hierarchy table  str abovetype  int abovex  str belowtype
int belowx
 "U" 0 "R" 1
 "R" 1 "A" 1
 "R" 1 "A" 2
 "R" 1 "A" 3
 "R" 1 "A" 4
 "R" 1 "A" 5
 "R" 1 "A" 6
 "R" 1 "A" 7
 "R" 1 "A" 8
 "R" 1 "A" 9
 "R" 1 "A" 10
 "R" 1 "A" 11
 "R" 1 "A" 12
 "R" 1 "A" 13
 "R" 1 "A" 14
 "R" 1 "A" 15
 "R" 1 "A" 16
 "R" 1 "A" 17
 "R" 1 "A" 18
 "R" 1 "A" 19
 "R" 1 "A" 20
 "R" 1 "A" 21
 "R" 1 "A" 22
 "R" 1 "A" 23
 "R" 1 "A" 24
 "R" 1 "A" 25
 "R" 1 "A" 26
 "R" 1 "A" 27
 "R" 1 "A" 28
!entry.M2L.unit.name single str
```

```
 ""
!entry.M2L.unit.positions table  dbl x  dbl y  dbl z
 -1.149000 3.109000 1.199000
 -1.741000 2.328000 0.746000
 -1.482000 3.597000 2.534000
 -1.765000 4.669000 2.478000
 -0.284000 3.440000 3.441000
 0.070000 4.381000 4.167000
 -2.669000 2.779000 3.102000
 -2.387000 1.707000 3.160000
 -3.550000 2.894000 2.437000
 -3.023000 3.296000 4.520000
 -3.306000 4.368000 4.464000
 -2.143000 3.181000 5.187000
 -4.195000 2.489000 5.081000
 -3.913000 1.417000 5.139000
 -5.076000 2.603000 4.415000
 -4.544000 2.999000 6.480000
 -4.827000 4.071000 6.424000
 -3.664000 2.885000 7.147000
 -5.671000 2.236000 7.093000
 -5.409000 1.200000 7.149000
 -6.943000 2.381000 6.340000
 -7.732000 1.812000 6.853000
 -6.831000 1.996000 5.318000
 -7.226000 3.443000 6.304000
 -5.894000 2.755000 8.447000
 -6.709000 2.192000 8.926000
 -6.167000 3.819000 8.390000
 -4.973000 2.643000 9.039000
!entry.M2L.unit.residueconnect table  int c1x  int c2x  int c3x  int c4x
int c5x  int c6x
 0 0 0 0 0 0
!entry.M2L.unit.residues table  str name  int seq  int childseq  int
startatomx  str restype  int imagingx
 "M2L" 1 29 1 "?" 0
!entry.M2L.unit.residuesPdbSequenceNumber array int
 1
!entry.M2L.unit.solventcap array dbl
 -1.000000
 0.0
 0.0
 0.0
 0.0
!entry.M2L.unit.velocities table  dbl x  dbl y  dbl z
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
```

```
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
```

A.2 AMBER Library File for Trimethyllysine

!!index array str

```
 "M3L"
!entry.M3L.unit.atoms table  str name  str type  int typex  int resx  int
flags  int seq  int elmnt  dbl chg
 "N" "" 0 1 131072 1 7 0.0
 "H" "" 0 1 131072 2 1 0.0
 "CA" "" 0 1 131072 3 6 0.0
 "HA" "" 0 1 131072 4 1 0.0
 "C" "" 0 1 131072 5 6 0.0
 "O" "" 0 1 131072 6 8 0.0
 "CB" "" 0 1 131072 7 6 0.0
 "HB1" "" 0 1 131072 8 1 0.0
 "HB2" "" 0 1 131072 9 1 0.0
 "CG" "" 0 1 131072 10 6 0.0
 "HG1" "" 0 1 131072 11 1 0.0
 "HG2" "" 0 1 131072 12 1 0.0
 "CD" "" 0 1 131072 13 6 0.0
 "HD1" "" 0 1 131072 14 1 0.0
 "HD2" "" 0 1 131072 15 1 0.0
 "CE" "" 0 1 131072 16 6 0.0
 "HE1" "" 0 1 131072 17 1 0.0
 "HE2" "" 0 1 131072 18 1 0.0
 "NZ" "" 0 1 131072 19 7 0.0
 "CZ1" "" 0 1 131072 20 6 0.0
 "HZ11" "" 0 1 131072 21 1 0.0
 "HZ12" "" 0 1 131072 22 1 0.0
 "HZ13" "" 0 1 131072 23 1 0.0
 "CZ2" "" 0 1 131072 24 6 0.0
 "HZ21" "" 0 1 131072 25 1 0.0
 "HZ22" "" 0 1 131072 26 1 0.0
 "HZ23" "" 0 1 131072 27 1 0.0
 "CZ3" "" 0 1 131072 28 6 0.0
 "HZ31" "" 0 1 131072 29 1 0.0
 "HZ32" "" 0 1 131072 30 1 0.0
 "HZ33" "" 0 1 131072 31 1 0.0
!entry.M3L.unit.atomspertinfo table  str pname  str ptype  int ptypex
int pelmnt  dbl pchg
 "N" "" 0 -1 0.0
 "H" "" 0 -1 0.0
 "CA" "" 0 -1 0.0
 "HA" "" 0 -1 0.0
 "C" "" 0 -1 0.0
 "O" "" 0 -1 0.0
 "CB" "" 0 -1 0.0
```

```
 "HB1" "" 0 -1 0.0
 "HB2" "" 0 -1 0.0
 "CG" "" 0 -1 0.0
 "HG1" "" 0 -1 0.0
 "HG2" "" 0 -1 0.0
 "CD" "" 0 -1 0.0
 "HD1" "" 0 -1 0.0
 "HD2" "" 0 -1 0.0
 "CE" "" 0 -1 0.0
 "HE1" "" 0 -1 0.0
 "HE2" "" 0 -1 0.0
 "NZ" "" 0 -1 0.0
 "CZ1" "" 0 -1 0.0
 "HZ11" "" 0 -1 0.0
 "HZ12" "" 0 -1 0.0
 "HZ13" "" 0 -1 0.0
 "CZ2" "" 0 -1 0.0
 "HZ21" "" 0 -1 0.0
 "HZ22" "" 0 -1 0.0
 "HZ23" "" 0 -1 0.0
 "CZ3" "" 0 -1 0.0
 "HZ31" "" 0 -1 0.0
 "HZ32" "" 0 -1 0.0
 "HZ33" "" 0 -1 0.0
!entry.M3L.unit.boundbox array dbl
 -1.000000
 0.0
 0.0
 0.0
 0.0
!entry.M3L.unit.childsequence single int
 2
!entry.M3L.unit.connect array int
 0
 0
!entry.M3L.unit.connectivity table  int atom1x  int atom2x  int flags
 1 3 1
 1 2 1
 3 7 1
 3 4 1
 3 5 1
 5 6 1
 7 10 1
 7 8 1
 7 9 1
 10 13 1
 10 12 1
 10 11 1
```

```
 13 16 1
 13 14 1
 13 15 1
 16 19 1
 16 18 1
 16 17 1
 19 20 1
 19 24 1
 19 28 1
 20 22 1
 20 21 1
 20 23 1
 24 27 1
 24 25 1
 24 26 1
 28 31 1
 28 30 1
 28 29 1
!entry.M3L.unit.hierarchy table  str abovetype  int abovex  str belowtype
int belowx
 "U" 0 "R" 1
 "R" 1 "A" 1
 "R" 1 "A" 2
 "R" 1 "A" 3
 "R" 1 "A" 4
 "R" 1 "A" 5
 "R" 1 "A" 6
 "R" 1 "A" 7
 "R" 1 "A" 8
 "R" 1 "A" 9
 "R" 1 "A" 10
 "R" 1 "A" 11
 "R" 1 "A" 12
 "R" 1 "A" 13
 "R" 1 "A" 14
 "R" 1 "A" 15
 "R" 1 "A" 16
 "R" 1 "A" 17
 "R" 1 "A" 18
 "R" 1 "A" 19
 "R" 1 "A" 20
 "R" 1 "A" 21
 "R" 1 "A" 22
 "R" 1 "A" 23
 "R" 1 "A" 24
 "R" 1 "A" 25
 "R" 1 "A" 26
 "R" 1 "A" 27
```

```
 "R" 1 "A" 28
 "R" 1 "A" 29
 "R" 1 "A" 30
 "R" 1 "A" 31
!entry.M3L.unit.name single str
 ""
!entry.M3L.unit.positions table  dbl x  dbl y  dbl z
 -1.149000 3.109000 1.199000
 -1.762000 2.473000 0.738000
 -1.482000 3.597000 2.534000
 -1.765000 4.662000 2.499000
 -0.284000 3.440000 3.441000
 -0.029000 4.327000 4.253000
 -2.654000 2.759000 3.062000
 -2.354000 1.701000 3.091000
 -3.515000 2.881000 2.385000
 -3.038000 3.236000 4.495000
 -3.319000 4.301000 4.454000
 -2.176000 3.112000 5.166000
 -4.228000 2.409000 5.044000
 -3.946000 1.347000 5.070000
 -5.084000 2.543000 4.376000
 -4.580000 2.914000 6.503000
 -4.828000 3.987000 6.444000
 -3.684000 2.788000 7.133000
 -5.751000 2.192000 7.181000
 -7.010000 2.360000 6.406000
 -7.835000 1.856000 6.934000
 -6.901000 1.916000 5.404000
 -7.243000 3.432000 6.307000
 -5.945000 2.785000 8.518000
 -6.783000 2.282000 9.026000
 -6.171000 3.858000 8.417000
 -5.028000 2.659000 9.115000
 -5.467000 0.740000 7.346000
 -6.314000 0.259000 7.862000
 -4.554000 0.609000 7.948000
 -5.329000 0.267000 6.363000
!entry.M3L.unit.residueconnect table  int c1x  int c2x  int c3x  int c4x
int c5x  int c6x
 0 0 0 0 0 0
!entry.M3L.unit.residues table  str name  int seq  int childseq  int
startatomx  str restype  int imagingx
 "M2L" 1 33 1 "?" 0
!entry.M3L.unit.residuesPdbSequenceNumber array int
 1
!entry.M3L.unit.solventcap array dbl
 -1.000000
```

```
 0.0
 0.0
 0.0
 0.0
!entry.M3L.unit.velocities table  dbl x  dbl y  dbl z
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
```

B.1 Hydrogen Bonds Analysis

Table B.1.1: Hydrogen bonds formed between JMJD2A-tudor and the histone peptides

Structure	Donor	AcceptorH	Acceptor	%Occupied	Distance (Å)	Angle (Degrees)
Tudor – H3K4me3	L-Thr3:O	R-Ser938:HG	R-Ser938:OG	97,69	2,68	17,25
	R-Asp945:OD1	L-Arg2:HH21	L-Arg2:NH2	45,7	2,83	26,94
	R-Asp945:OD2	L-Arg2:HH21	L-Arg2:NH2	44,99	2,83	26,94
	R-Asp945:OD1	L-Arg2:HE	L-Arg2:NE	35,39	2,85	24,34
	R-Asp945:OD2	L-Arg2:HE	L-Arg2:NE	34,43	2,85	24,44
	L-Thr3:O	R-Asn940:HD21	R-Asn940:ND2	12,78	2,94	35,19
	R-Glu944:OE1	L-Arg2:HH22	L-Arg2:NH2	12,17	2,83	27,38
	R-Asp939:OD1	L-Gln5:HE21	L-Gln5:NE2	11,86	2,85	20,47
	R-Glu944:OE2	L-Arg2:HH22	L-Arg2:NH2	9,58	2,82	27,24
Tudor – H4K20me3	R-Glu929:OE2	L-Arg17:HH22	L-Arg17:NH2	43,08	2,8	27,38
	R-Glu929:OE2	L-Arg17:HH12	L-Arg17:NH1	39,14	2,79	27,29
	L-Leu22:O	R-Trp967:HE1	R-Trp967:NE1	35,16	2,84	26,76
	R-Glu929:OE1	L-Arg17:HH12	L-Arg17:NH1	29,16	2,8	24,95
	R-Glu929:OE1	L-Arg17:HH22	L-Arg17:NH2	27,96	2,81	27,13
	R-Asp939:OD1	L-Arg19:HH12	L-Arg19:NH1	26,16	2,81	23,75
	L-Arg23:OXT	R-Thr968:HG1	R-Thr968:OG1	21,22	2,66	14,98
	R-Asp939:OD2	L-Arg19:HH21	L-Arg19:NH2	18,02	2,84	25,92
	R-Asp939:OD2	L-Arg19:HH22	L-Arg19:NH2	17,69	2,82	24,73
	L-Arg23:O	R-Thr968:HG1	R-Thr968:OG1	15,61	2,68	15,53
	R-Asp939:OD2	L-Arg19:HH12	L-Arg19:NH1	15,35	2,82	25,55
	R-Asp939:OD1	L-Arg19:HH22	L-Arg19:NH2	15,25	2,83	27,02
	R-Asp939:OD1	L-Arg19:HE	L-Arg19:NE	15,22	2,84	23,21
	R-Asp939:OD1	L-Arg19:HH21	L-Arg19:NH2	14,73	2,84	26,02
	R-Asp939:OD2	L-Arg19:HE	L-Arg19:NE	13,49	2,84	24,79

	L-Arg23:OXT	R-Trp967:HE1	R-Trp967:NE1	12,41	2,82	21,33
	R-Ser936:O	L-Arg17:HE	L-Arg17:NE	10,04	2,84	24,58
	R-Thr968:O	L-Arg23:HE	L-Arg23:NE	6,84	2,85	32,32
	R-Thr968:O	L-Arg23:HH21	L-Arg23:NH2	6,31	2,84	34,64
	L-Arg23:O	R-Trp967:HE1	R-Trp967:NE1	5,84	2,83	20,65
	R-Phe932:O	L-Arg17:HH21	L-Arg17:NH2	5,8	2,83	37,42
Tudor – H4K20me2	L-Arg23:O	R-Trp967:HE1	R-Trp967:NE1	76,9	2,82	18,4
	R-Ser936:O	L-Arg17:HH11	L-Arg17:NH1	22,46	2,84	24,97
	R-Phe932:O	L-Arg17:HH12	L-Arg17:NH1	20,57	2,82	36,61
	R-Asp969:OD1	L-Arg23:HH11	L-Arg23:NH1	19,73	2,82	22,97
	R-Glu929:OE2	L-Arg17:HH12	L-Arg17:NH1	16,67	2,8	27,28
	R-Glu929:OE2	L-Arg17:HH22	L-Arg17:NH2	15,98	2,79	28,43
	R-Asp939:OD2	L-Arg19:HH12	L-Arg19:NH1	13,63	2,82	25,94
	R-Asp939:OD1	L-Arg19:HH21	L-Arg19:NH2	11,46	2,84	26,42
	R-Asp939:OD1	L-Arg19:HH12	L-Arg19:NH1	11,25	2,81	24,07
	R-Asp939:OD2	L-Arg19:HH21	L-Arg19:NH2	10,9	2,84	26,45
	R-Asp939:OD2	L-Arg19:HE	L-Arg19:NE	10,16	2,84	24,03
	R-Asp939:OD2	L-Arg19:HH22	L-Arg19:NH2	10,14	2,83	27,19
	R-Asp934:OD2	L-K20me2:HZ	L-K20me2:NZ	9,53	2,82	32,04
	R-Asp939:OD1	L-Arg19:HE	L-Arg19:NE	8,79	2,83	24,52
	R-Ser936:O	L-Arg17:HE	L-Arg17:NE	8,6	2,84	25,6
	R-Asp969:OD2	L-Arg23:HH11	L-Arg23:NH1	8,42	2,82	24,47
	R-Asp939:OD1	L-Arg19:HH22	L-Arg19:NH2	8,03	2,83	26,03
	R-Glu929:OE1	L-Arg17:HH12	L-Arg17:NH1	6,32	2,79	26
	R-Glu929:OE1	L-Arg17:HH22	L-Arg17:NH2	6,06	2,8	29,01
	R-Asp934:OD1	L-K20me2:HZ	L-K20me2:NZ	6,02	2,81	29,61
	R-Ser936:O	L-Arg17:HH21	L-Arg17:NH2	5,19	2,84	32,01

Tudor – H3K9me3	L-Arg8:O	R-Ser938:HG	R-Ser938:OG	66,57	2,7	16,96
	R-Asn940:OD1	L-Arg8:HE	L-Arg8:NE	28,53	2,85	22,96
	R-Thr968:OG1	L-Arg8:HH11	L-Arg8:NH1	26	2,88	27,11
	L-Arg8:O	R-Asn940:HD21	R-Asn940:ND2	25,1	2,88	25,19
	R-Asn940:O	L-Arg8:HH21	L-Arg8:NH2	16,39	2,85	31,78
	R-Asp1001:OD2	L-Ala7:H2	L-Ala2:N	8,25	2,8	33,35
	R-Asn940:OD1	L-Arg8:HH21	L-Arg8:NH2	7,83	2,87	34,79
	R-Asp1001:OD2	L-Ala7:H1	L-Ala7:N	6,85	2,81	34,53
	R-Asp1001:OD2	L-Ala7:H3	L-Ala7:N	5,29	2,81	32,53

B.2 Salt Bridges Analysis

Table B.2.1: Salt bridges formed between JMJD2A-tudor and the histone peptides

Complex	Donor	Acceptor	%Occupied	Distance (Å)
Tudor – H3K4me3	R-Asp945:OD1	L-Arg2:NH2	96,43	3,11
	R-Asp945:OD2	L-Arg2:NH2	94,61	3,11
	R-Asp945:OD1	L-Arg2:NE	91,85	3,18
	R-Asp945:OD2	L-Arg2:NE	88,38	3,19
	R-Glu944:OE1	L-Arg2:NH2	27,12	3,18
	R-Glu944:OE2	L-Arg2:NH2	22,22	3,21
	R-Asp934:OD1	L-K4me3:NZ	6,13	3,82
	R-Glu944:OE1	L-Arg2:NH1	5,29	3,82
	R-Asp934:OD2	L-K4me3:NZ	4,57	3,82
	R-Glu944:OE2	L-Arg2:NH1	4,13	3,81
Tudor – H4K20me3	R-Asp939:OD2	L-Arg19:NH2	72,12	3,12
	R-Asp939:OD1	L-Arg19:NH2	70,98	3,18
	R-Glu929:OE2	L-Arg17:NH2	62,67	2,98
	R-Glu929:OE1	L-Arg17:NH2	61,49	3,18
	R-Glu929:OE2	L-Arg17:NH1	51,94	2,95
	R-Asp939:OD1	L-Arg19:NH1	45,75	3,04
	R-Glu929:OE1	L-Arg17:NH1	40,86	3
	R-Asp939:OD2	L-Arg19:NH1	38,94	3,19
	R-Asp939:OD2	L-Arg19:NE	28,88	3,19
	R-Asp939:OD1	L-Arg19:NE	28,76	3,13
	R-Asp969:OD2	L-Arg23:NH2	4,29	3,11
	R-Asp969:OD2	L-Arg23:NE	4,18	3,11

Tudor – H4K20me2	R-Asp939:OD2	L-Arg19:NH2	43,61	3,13
	R-Asp939:OD1	L-Arg19:NH2	42,24	3,15
	R-Asp969:OD1	L-Arg23:NH1	27,49	2,95
	R-Asp939:OD2	L-Arg19:NH1	23,12	3,06
	R-Glu929:OE2	L-Arg17:NH2	22,2	2,96
	R-Glu929:OE2	L-Arg17:NH1	19,94	2,87
	R-Asp934:OD2	L-K20me2:NZ	19,79	3,19
	R-Asp939:OD1	L-Arg19:NH1	19,53	3,07
	R-Asp939:OD2	L-Arg19:NE	19,06	3,13
	R-Asp939:OD1	L-Arg19:NE	18,01	3,16
	R-Asp969:OD1	L-Arg23:NE	15,83	3,63
	R-Asp934:OD1	L-K20me2:NZ	12,96	3,23
	R-Asp969:OD2	L-Arg23:NH1	12,69	2,99
	R-Asp929:OE1	L-Arg17:NH2	11,6	3,16
	R-Asp969:OD2	L-Arg23:NE	8,13	3,61
	R-Asp929:OE1	L-Arg17:NH1	7,93	2,92
	R-Asp969:OD1	L-Arg23:NH2	4,7	3,13
	R-Asp969:OD2	L-Arg23:NH2	3,72	3,11
Tudor – H3K9me3	R-Asp934:OD2	L-K9me3:NZ	10,04	3,81
	R-Asp934:OD1	L-K9me3:NZ	9,7	3,81
	R-Asp945:OD2	L-Arg8:NH2	8,53	3,31
	R-Asp945:OD1	L-Arg8:NH2	7,07	3,16
	R-Asp945:OD1	L-Arg8:NE	2,04	3,59

BIBLIOGRAPHY

[1] H. Tan, S. Wu, J. Wang and Z.K. Zhao, The JMJD2 members of histone demethylase revisited, *Mol. Biol. Rep.* **35** (2007), pp. 551-556.

[2] J. Kim, J. Daniel, A. Espejo, A. Lake, M. Krishna, L. Xia, Y. Zhang and M.T. Bedford, Tudor, MBT and chromo domains gauge the degree of lysine methylation, *EMBO Rep.* **7** (2006), pp. 397–403.

[3] J. Lee, J.R. Thompson, M.V. Botuyan, G. Mer, Distinct binding modes specify the recognition of methylated histones H3K4 and H4K20 by JMJD2A-tudor. *Nat. Struct. Mol. Biol.* **15** (2008), pp.109–111.

[4] Y. Zhang, It takes a PHD to interpret histone methylation, *Nat. Struct. Mol. Biol.* **13** (2006), pp. 572–574.

[5] M.S. Torok, P.A. Grant, The generation and recognition of histone methylation, *Results Probl. Cell. Differ* **41** (2006) pp. 25-46.

[6] Y. Zhang, D. Reinberg, Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails, *Genes Dev.* **15** (2001), pp. 2343–2360.

[7] A. Barski *et al.*, High-resolution profiling of histone methylations in the human genome, *Cell* **129** (2007), pp. 823–837.

[8] D. Zhang, H.G. Yoon and J. Wong, JMJD2A is a novel N-CoR-interacting protein and is involved in repression of the human transcription factor achaete scute-like homologue 2 (ASCL2/Hash2), *Mol. Cell. Biol.* **25** (2005), pp. 6404–6414.

[9] M. S. Cosgrove, PHinDing a New Histone "Effector" Domain, *Structure* **14** (2006), pp.1096-1098.

[10] X. de la Cruz, S. Lois, S. Sanchez-Molina and M.A. Martinez-Balbas, Do protein motifs read the histone code?, *Bioessays* **27** (2005), pp. 164–175.

[11] C. L. Peterson, M.A. Laniel, Histones and histone modifications, *Curr. Biol.* **14** (2004), pp. R546–R551.

[12] Z. Chen, J. Zang, J. Kappler, X. Hong, F. Crawford, Q. Wang, F. Lan, C. Jiang, J. Whetstine and S. Dai *et al.*, Structural basis of the recognition of a methylated histone tail by JMJD2A, *Proc Natl Acad Sci USA* **104** (2007), pp. 10818–10823.

[13] L. Corsini and M. Sattler, Tudor hooks up with DNA repair, *Nat. Struct. Mol. Biol.* **14** (2007), pp. 98–99.

[14] J.K. Wiencke, S. Zheng, Z. Morrison, R-F. Yeh, Differentially expressed genes are marked by histone 3 lysine 9 trimethylation in human cancer cells, *Oncogene* **27** (2008); pp. 2412–21.

[15] J. R. Wilson, Targeting the JMJD2A histone lysine demethylase, *Nat Struct Mol Biol.* **14** (2007) 682–684.

[16] J.F. Couture, E. Collazo, P.A. Ortiz-Tello, J.S. Brunzelle and R.C. Trievel, Specificity and mechanism of JMJD2A, a trimethyllysine-specific histone demethylase, *Nat Struct Mol Biol* **14** (2007), pp. 689–695.

[17] E. Metzger, R. Schüle, The expanding world of histone lysine demethylases, *Nat Struct Mol Biol.* **14** (2007) pp. 252–254.

[18] S.G. Gray, A.H. Iglesias, F. Lizcano, R. Villanueva, S. Camelo, H. Jingu, B.T. Teh, N. Koibuchi, W.W. Chin, E. Kokkotou and F. Dangond, Functional characterization of JMJD2A, a histone deacetylase- and retinoblastoma-binding protein, *J. Biol. Chem.* **280** (2005), pp. 28507–28518.

[19] S. Shin, R. Janknecht, Activation of androgen receptor by histone demethylases JMJD2A and JMJD2D, *Biochem. Biophys. Res. Commun.* **359** (2007), pp. 742–746.

[20] G. Kustatscher, A. G. Ladurne, Modular paths to ‘decoding’ and ‘wiping’ histone lysine methylation, *Current Opinion in Chemical Biology* **11** (2007), pp. 628-635.

[21] Y. Huang, J. Fang, M.T. Bedford, Y. Zhang and R.M. Xu, Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A, *Science* **312** (2006), pp. 748–751.

[22] G. Charier *et al.* The Tudor tandem of 53BP1: A new structural motif involved in DNA and RG-rich peptide binding. *Structure (Camb.)* **12** (2004), pp. 1551–1562.

[23] T. K. Kundu, D. Dasgupta, Chromatin and disease, (2007), Springer, Berlin.

[24] K Ogata, H. Umeyama, An automatic homology modeling method consisting of database searches and simulated annealing, *J Mol Graph Model*, **18** (2000) pp. 258-272.

[25] K. Arnold, L. Bordoli, J. Kopp and T. Schwede,The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling, *Bioinformatics* **22** (2006), pp. 195–201.

[26] T. Schwede *et al*., SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res*, **31** (2003), pp. 3381–3385.

[27] R. Rodriguez, G. Chinea, N. Lopez, T. Pons, G. Vriend, Homology modeling, model and software evaluation: three related resources. *Bioinformatics* **14** (1998), pp. 523–528.

[28] Hillisch A, Pineda LF, Hilgenfeld R. Utility of homology models in the drug discovery process, *Drug Discov. Today*, **9** (2004), pp. 659–669.

[29] R. Sanchez, U. Pieper, F. Melo, N. Eswar, M.A. Marti-Renom, M.S. Madhusudhan, N. Mirkovic and A. Sali , Protein structure modeling for structural genomics. *Nat. Struct. Biol.* **7** (2000), pp. 986–990.

[30] S. Y. Chung, S. Subbiah, A structural explanation for the twilight zone of protein sequence homology. *Structure*, **4** (1996), pp. 1123–27.

[31] J.A. McCammon, B.R. Gelin, M. Karplus, Dynamics of folded proteins, *Nature*, **267** (1977), pp. 585−590.

[32] M. Karplus, J.A. McCammon, Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **9** (2002), pp. 646–652.

[33] A. Grossfield, S. E. Feller, M. C. Pitman, Convergence of molecular dynamics simulations of membrane proteins, *Proteins: Struct. Funct. Bioinf.*, **67** (2007), pp. 31−40.

[34] C. A. Sotriffer, W. Flader, R. H. Winger, B. M. Rode, K. R. Liedl, J. M. Varga, Automated docking of ligands to antibodies: Methods and applications, *Methods* **20** (2000), pp. 280-291.

[35] B. Ondul, Structure based drug design, MS Thesis, Koc University (2009).

[36] C. Hetenyi, D. van der Spoel, Efficient docking of peptides to proteins without prior knowledge of the binding site, *Protein Science* **11** (2002), pp. 1729-1737.

[37] A. Wolf, M. Zimmermann, M. Hofmann-Apitius, Alternative to consensus scoring--a new approach toward the qualitative combination of docking algorithms, J. Chem. Inf. Model 47 (2007), pp. 1036-1044.

[38] T. Schwede, J. Kopp, N. Guex, M. C. Peitsch, SWISS-MODEL: an automated protein homology-modeling server, *Nucleic Acids Research* **31** (2003), pp. 3381-3385.

[39] Y. Duan, C. Wu, S. Chowdhury, M.C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang and P. Kollman, A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations, *J. Comput. Chem.* **24** (2003), pp. 1999–2012.

[40] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, *J. Am. Chem. Soc.* **117** (1995), pp. 5179-5197.

[41] J. Wang, R.M. Wolf, J.W. Cardwell, P.A. Kollman, D.A. Case, Development and Testing of a General Amber Force Field, *Journal of Computational Chemistry* **25** (2004), pp. 1157–1174.

[42] C. I. Bayly, P. Cieplak, W. D. Cornell, P. A. Kollman, A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model, *J. Phys. Chem.* **97** (1993) pp. 10269–10280.

[43] Discovery Studio Visualizer, Accelrys Inc., San Diego, CA, USA.

[44] M.J.T. Frisch, G.W. Schlegel, H.B. Scuseria, G.E. Robb, M.A. Cheeseman and J.R. Montgomery *et al.*, *Gaussian 03*, Gaussian, Inc, Pittsburgh, PA (2003).

[45] A. Pigache, P. Cieplak & F.-Y. Dupradeau, Automatic and highly reproducible RESP and ESP charge derivation: Application to the development of programs RED and X RED, 227th ACS National Meeting, Anaheim, CA, USA, March 28 - April 1, 2004.

[46] D. A. Case, T. E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K. M. Merz Jr., A. Onufriev, C. Simmerling, B. Wang, R. J. Woods, The Amber biomolecular simulation programs, *J. Comput. Chem.* **26** (2005), pp. 1668-1688.

[47] M. D. Eldride, C. W. Murray, T. R. Auton, G. V. Paolini, R. P. Mee, Empirical scoring functions. I: The development of fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes, *J. Computer-Aided Mol. Des.* **11** (1997), pp 425-445.

[48] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, A. J. Olson, Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, Journal of Computational Chemistry 19 (1998), pp. 1639-1662.

[49] R. Huey, G. M. Morris, A. J. Olson, D. S. Goodsell, A semiempirical free energy force field with charge-based desolvation, *Journal of Computational Chemistry* **28** (2007), pp. 1145-1152.

[50] J. Gasteiger, M. Marsili, A new model for calculating atomic charges in molecules, *Tetrahedron Letters* **34** (1978), pp. 3181–3184.

335-340.

[51] D. A. Case, T.A. Darden, T.E. Cheatham III, C.L. Simmerling, J. Wang and R.E. Duke *et al.*, *AMBER 10*, University of California, San Francisco, CA (2008).

[52] W. L. Jorgensen, Transferable intermolecular potential functions for water, alcohols and ethers. Application to liquid water, *J. Am. Chem. Soc.***103** (1981), pp.

[53] L. Kale, R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, K. Schulten, NAMD2: Greater scalability for parallel molecular dynamics, *J. Comput. Phys.* **151** (1999), pp. 283–312.

[54] S. Miyamoto, P. A. Kollman, SETTLE: An analytical version of the SHAKE and RATTLE algorithm for rigid water models". *Journal of Computational Chemistry* **13** (1992), pp. 952–962.

[55] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, A smooth particle mesh Ewald method, *J. Chem. Phys.* **103** (1995), pp. 8577–9593.

[56] S. E. Feller, Y. Zhang, R.W. Pastor, B. R. Brooks, Constant pressure molecular dynamics simulation: The Langevin piston method, *J. Chem. Phys.* **103**(11), 1995, pp.4613–4621.

[57] G. J. Martyna, D. J. Tobias, M. L. Klein, Constant pressure molecular dynamics algorithms, *J. Chem. Phys.* **101**(5), 1994, pp. 4177-4189.

[58] W. Rocchia, S. Sridharan, A. Nicholls, E. Alexov, A. Chiabrera, B. Honig, Rapid grid-based construction of the molecular surface for both molecules and geometric objects: Applications to the finite difference Poisson-Boltzmann method, *J. Comput. Chem*. **23** (2002), pp. 128-137.

[59] D. Sitkoff, K. A. Sharp, B. Honig, Accurate calculation of hydration free energies using macroscopic solvent models, *J. Phys. Chem*. **98** (1994), pp. 1978-1988.

[60] A. Onufriev, D. Bashford, D. A. Case, Exploring protein native states and large-scale conformational changes with a modified generalized born model, *Proteins* **55** (2004) pp. 383–394.

[61] M.L. Connolly, Analytical molecular surface calculation, *J. Appl. Crystallogr.* **16** (1983), pp. 548-558.

[62] J. Weiser, P.S.W. Shenkin, W.K. Clark Still, Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO), *J. Comput. Chem.* **20** (1999), pp. 217-230.

[63] H. Gohlke, D. A. Case, Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf, *J.Comp. Chem.* **25** (2004), pp. 238-250.

[64] H. Gohlke, C. Kiel, D. A. Case, Insights into protein–protein binding by binding free energy calculation and free energy decomposition for the Ras–Raf and Ras–RalGDS complexes, J. Mol. Biol. **330** (2003), pp. 891-913.

[65] R. Zhou, G. Krilov, B.J. Berne, Comment on "Can a continuum solvent model reproduce the free energy landscape of a β-hairpin folding in water?" The Poisson-Boltzmann equation, J. Phys. Chem. B 108 (2004), pp. 7528-7530.

[66] I. Stoica, S. K. Sadiq, P. V. Coveney, Rapid and accurate prediction of binding free energies for saquinavir-bound HIV-1 proteases, *J. Am. Chem. Soc.* **130** (2008), pp. 2639-2648.

[67] W. Chen, C. E. Chang, M. K. Gilson, Calculation of Cyclodextrin binding affinities: Energy, entropy, and implications for drug design, *Biophys. J.* **87**, 2004, pp. 3035–3049.

[68] C. E. Chang, W. Chen, M. K. Gilson, Ligand configurational entropy and protein binding, *Proc. Natl. Acad. Sci. U.S.A.* **104** (2007), pp. 1534-1539.

VITA

Musa Özboyacı was born in Diyarbakır, Turkey in 1985. He received his Bachelor of Science Degree in Molecular Biology and Genetics from Istanbul Technical University. In September 2007 he attended Computational Sciences and Engineering M.S. program at Koç University. From 2007 to 2009, he worked as research and teaching assistant at the same institution.