# DETERMINATION OF STATISTICAL CORRELATIONS BETWEEN METHYLATED LYSINE RESIDUES OF HISTONE H3 TAIL BY MOLECULAR DYNAMICS SIMULATIONS

**by**

**Deniz ŞANLI**

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

**Master of Science**
**in**
**Chemical & Biological Engineering**

**Koc University**

**September 2009**

Koc University

Graduate School of Sciences and Engineering

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

Deniz Şanlı

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:

_____

Assc. Prof. Özlem KESKİN (Advisor)

_____

Prof. Burak ERMAN

_____

Assc. Prof. Attila GÜRSOY

Date: _____

## ABSTRACT

Post-translational modifications of histone tails have been a very favored topic in Molecular Biology since the proposal of "histone code" hyphothesis. Histone H3 tail is the most well-known one of the histone proteins that undergoes several chemical modifications thus affecting a variety of physiological processes. Moreover, an expanding knowledge of communications between different modifications have emerged in recent years. Although the downstream outcomes of the most of these modifications have been revealed, the detailed mechanism of this dynamic cross-talk between different modifications still remain to be elucidated. In this study, the cooperativity between histone H3 lysine 4 (H3K4) and lysine 9 (H3K9) methylations were aimed to be investigated by Molecular Dynamics (MD) simulations. For this purpose, MD simulations of three types of H3 tails (unmodified H3 tail, tri-methylated K4 (H3K4me3), tri-methylated K9 (H3K9me3)) were carried out. Subsequently, Modal Analysis was performed to the trajectories in order to reveal the distance fluctuation correlations between the residues. By using these residue correlations, statistical mechanical formulations were derived to relate the cooperativity between the methylated residues to the distance fluctuations at residue level. According to the correlation function results, the cross-regulation patterns between the modifications were predicted on statistical thermodynamics basis.

# ÖZET

Histon proteinlerinin post-translasyonel modifikasyonları, "histon kodu" hipotezinin ortaya atılmasından itibaren Moleküler Biyolojinin en önemli konularından biri haline gelmiştir. Histon H3 kuyruğu, histon proteinlerinin en çok çalışılanı olup, birçok kimyasal modifikasyona uğrayarak farklı fizyolojik prosesleri etkilediği bilinmektedir. Bununla birlikte, son yıllarda yapılan çalışmalarla farklı modifikasyonlar arasında iletişim olduğu ortaya çıkmıştır. Bu modifikasyonların birçoğunun hücre içi etkilerinin bilinmesine rağmen, farklı modifikasyonlar arasındaki bu dinamik iletişimin detaylı mekanizması çözümlenmeyi beklemektedir. Bu çalışmada, histone H3 lizin 4 (H3K4) ve histone H3 lizin 9 (H3K9) metilasyonları arasındaki cooperativitenin, Moleküler Dinamik (MD) simülasyonları ile incelenmesi amaçlanmıştır. Bu amaçla, üç farklı histon H3 kuyruğunun (modifiye edilmemiş H3 kuyruğu, üç metil grubu eklenmiş K4 (H3K4me3), üç metil grubu eklenmiş K9 (H3K9me3)) MD simulasyonları gerçekleştirilmiştir. Bunu takiben, amino asitler arasındaki uzaklık değişimi korelasyonlarının araştırılması amacıyla Mod Analizi uygulanmıştır. Bu korelasyonlar yardımıyla, cooperativiteyi uzaklık değişimleri ile ilişkilendirmek amacıyla istatistiksel mekanik formülasyonları elde edilmiştir. Korelasyon fonksiyonu sonuçlarına göre modifikasyonlar arasındaki kontrol mekanizması, istatistiksel temodinamik bazında belirlenmiştir.

**ACKNOWLEDGEMENTS**

First of all I would like to thank my advisors Prof. Burak Erman, Assoc. Prof. Özlem Keskin and Assoc. Prof Attila Gürsoy for their guidance, help and support throughout my studies.

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

I would like to thank my home mates Nazire Seda Yaşar, Serra Caner, my "histone mates" Musa Özboyacı and Özlem Ulucan. Also I would like to thank all of my friends at Koç University, Mehmet Çağrı Dedeoğlu, Turan Bulmuş, Nihan Çömden, Hakan Doğan, Özge Narin. Life at Yeni Uyum would be a nightmare without you.

I owe my deepest appreciation to my family, always supporting, loving unconditionally and being there for me. I wouldn't be breathing without you.

**TABLE OF CONTENTS**

## LIST OF TABLES

## LIST OF FIGURES

# Chapter 1

# INTRODUCTION

Nucleosomes are the fundamental structures of eukaryotic chromatin. Histone proteins are the basic components of nucleosomes which compact 1.8 m of DNA into eukaryotic nucleus. Each nucleosome core particle consists of four histone proteins, histone H2A, H2B, H3 and H4, and has an octameric structure formed by a central heterotetramer of histones H3 and H4, and two heterodimers of histones H2A and H2B. Additionally a linker protein, histone H1/H5, performs locking DNA on to the core domain. All together, they constitute a spoon-like structure for 146 bp of DNA to tangle around. With 10-60 bp of `linker' DNA mediating the connections between the nucleosomes, a 10 nm fiber structure, also named as `beads on a string', is formed which further generates a more condensed structure of 30 nm thick fiber. These thick fibers are the primary building blocks of in vivo 100-400 nm thicker interphase fibers or more compact metaphase chromosome structures [1].

Having such a basic structural role, histone proteins have N- and/or C-terminal tails that protrude from the core surface and undergo several post-translational modifications. These post-translational modifications of histone tails make them key switch points for chromatin condensation/decondensation, affecting the cellular downstream processes such as transcription and DNA repair [2].

Histone tails can undergo acetylation, methylation, phosphorylation, ubiquitination, sumoylation, biotinylation, citrullination, ADP-ribosylation, cis-trans isomerization. These

different types of modifications target specific amino acids. For instance, lysine residues (K) can undergo acetylation and methylation, arginine residues (R) can be subjected to methylation, citrullination, ADP-ribosylation, whereas serine (S) and threonine (T) residues are amenable to phosphorylation. Furthermore, these diverse modifications prefer the aforementioned residues at a specific tail and position. The lysine residue at position 4 of the histone H3 tail (H3K4) can undergo methylation and biotinylation, the lysine residue at position 9 of the same tail (H3K9) is subjected to methylation, acetylation and biotinylation, whereas the lysine residue at position 14 of the same tail (H3K14) is exposed to acetylation. In the case of methylation, the degree of modification can also be diverse. For instance, while the lysine residues can undergo mono-, di- or tri-methylation, the arginine residues can be subjected to mono-methylation, symmetric di-methylation and asymmetric di-methylation [3]. Specifically, H3K4 and H3K9 residues can be mono-, di- and tri-methylated, whereas the lysine 27 residue of the same tail (H3K27) can be mono- and di-methylated, but not tri-methylated. Each of these diverse modifications have different functional outcomes e.g. acetylation is generally associated with transcriptional activation whereas methylation can be associated with both activation and repression, depending on the residue that is exposed to the modification or the state of modification [3].

Among the unobstructed dynamic histone tails, H3 N-terminal tail is the most extensively studied and well-known tail so far. It consists of 40 amino acids projecting from the H3 core domain, and there are 12 residues proven to be chemically modified. These residues are R2, K4, K9, S10, K14, R17, K18, K23, R26, K27, S28 and K36, and the modification types vary from methylation to phosphorylation. However, among these 12 residues, K4, K9 and K14 are the ones, the modifications of which have the most significant outcomes. H3K4 can be methylated, H3K9 can be methylated and acetylated, whereas H3K14 can only be acetylated. As mentioned previously, each of these

modification types as well as their states have different functional outcomes. Acetylation of H3K9 and H3K14 forms transcriptionally active chromatin, euchromatin. Keeping in mind that the histone methylation has a dual role in regulation; H3K4 methylation is generally associated with transcriptional activation, whereas H3K9 methylation is correlated with inactive chromatin, also known as heterochromatin [4].

Recent studies reveal the dynamic interplay between different types of modifications on different residues, even between intra- and inter-nucleosomes, which was first proposed in the `histone code' hypothesis by Strahl and Allis in 2000 [5]. According to this hypothesis: "various modifications that occur on one or more histone termini act in a combinatorial and sequence-dependent manner to yield specific downstream events; these modifications are therefore used as a cellular vocabulary for the regulation of different transcription-based processes" [6]. Up to now, this dynamic cross-talk between the histone modifications has been stated to be pivotal as it harbors critical epigenetic information and the epigenetic mechanisms are essential during development, thus misregulation causes several diseases including cancer [2].

A remarkable outcome of the histone code hyphothesis is the significant cross-regulation patterns between H3K4, H3K9 and H3K14. The modifications of these residues can drastically promote or prevent each other. Addition of methyl groups on H3K4 prevents methylation of H3K9, whereas acetylation of H3K14 is promoted. Moreover, methylation of H3K9 prevents both methylation of H3K4 and acetylation of H3K14 [7]. Each of these epigenetic cross-regulations involves diverse downstream pathways in cellular processes.

Inspired by the dynamic cross-talk of the H3 tail modifications, the cooperative act of the aforementioned modifications of the three residues (H3K4, H3K9, H3K14) were aimed to be investigated. For this purpose, Molecular Dynamics simulations of the three types of histone H3 tails were carried out to account for the time dependent behavior. Subsequently,

Modal Analysis was performed for the observation of the role of fluctuation dynamics and residue correlations. And lastly, the statistical thermodynamic facts, such as the correlation function and the residue interaction energy, were studied for better understanding of the cooperativity mechanisms in theoretical base.

# Chapter 2

## THERMODYNAMICS OF FLUCTUATIONS

The second postulate of thermodynamics states that: "There exists a function (called the entropy S) of the extensive parameters of any composite system, defined for all equilibrium states and having the following property: The values assumed by the extensive parameters in the absence of an internal constraint are those that maximize the entropy over the manifold of constrained equilibrium states" [8]. Thus, similar to the 2nd law of thermodynamics, it implies that the entropy of isolated systems should aim to rise until the system equilibrium is attained. However, following the discovery of statistical mechanics, it became obvious that the 2nd law declares only a statistical consequence and there should be a nonzero probability that the entropy of an isolated system will decrease. That was the point the fluctuation theorem appears on the scene, expressing this nonzero probability precisely.

A thermodynamic system can be defined by its extensive parameters, the entropy (S), the internal energy (U), the volume (V), and the number of particles (N), which form the parameters of the fundamental relation of that system, represented in Eqn. (2.1) [8].

$$S = S(U, V, N) \tag{2.1}$$

These extensive parameters of a thermodynamic system persistently undergo fluctuations around their equilibrium values, which is also termed as the random transitions among the microstates of the system [8].

Analogous to Eqn. (2.1) a system that is in contact with a temperature (T), pressure (P) and a force (F) reservoir can be characterized with the following fundamental relation:

$$S = S(U, V, R) \tag{2.2}$$

where, U, V are the mean values of the energy and volume of the system respectively and **R** is a vector quantity representing the mean positions of $C^{\alpha}$ atoms of the system [9]. Due to the interchanges of the system with its surrounding reservoir, the aforementioned thermodynamic variables exhibit fluctuations around their equilibrium values.

For a system that is described with the fundamental relation given in Eqn. (2.2), the probability distribution with the instantaneous variables $\hat{U}, \hat{V}, \hat{R}$ can be expressed as:

$$f(\hat{U}, \hat{V}, \hat{R}) = \exp\left\{-k_B^{-1} S\left[\frac{1}{T}, \frac{P}{T}, \frac{F}{T}\right] - k_B^{-1}\left(\frac{1}{T}\hat{U} + \frac{P}{T}\hat{V} - \frac{F}{T}\hat{R}\right)\right\} \tag{2.3}$$

where, $k_B$ is the Boltzmann constant, $S\left[\frac{1}{T}, \frac{P}{T}, \frac{F}{T}\right]$ is the Massieu function of the entropy, T, P, **F** representing the temperature, pressure and force respectively and $\frac{1}{T}, \frac{P}{T}, \frac{F}{T}$ are the intensive parameters with the values that are equal to the reservoir [8].

The Massieu function of the entropy can be written explicitly as stated in Eqn. (2.4) [9].

$$S\left[\frac{1}{T},\frac{P}{T},\frac{F}{T}\right]=S-\frac{1}{T}U-\frac{P}{T}V+\frac{F}{T}R \tag{2.4}$$

If the expressions given in Eqn. (2.3) and (2.4) are combined, the probability distribution of the system can be obtained explicitly as given in Eqn. (2.5) below:

$$f(\hat{U},\hat{V},\hat{R})=\exp\left\{-k_B^{-1}\left(S-\frac{1}{T}U-\frac{P}{T}V+\frac{F}{T}R\right)-k_B^{-1}\left(\frac{1}{T}\hat{U}+\frac{P}{T}\hat{V}-\frac{F}{T}\hat{R}\right)\right\} \tag{2.5}$$

Following Eqn. (2.5), the correlation of two fluctuating variables can be derived as:

$$\left\langle \Delta\hat{X}_i\Delta\hat{X}_j\right\rangle \equiv \sum\left(\hat{X}_i-X_i\right)\left(\hat{X}_j-X_j\right)f\left(\hat{U},\hat{V},\hat{R}\right) \tag{2.6}$$

where, X indicates any of the extensive variables U, V, **R** and $\hat{X}$ represents the instantaneous values of the aforementioned variables, thus being any of $\hat{U},\hat{V},\hat{R}$ [9]. The subscripts i and j stands for the residue index. The $\left\langle \Delta\hat{X}_i\Delta\hat{X}_j\right\rangle$ term is called the second central moment of the fluctuations of the variable $\hat{X}$ which measures the correlation of the fluctuations of the two variables $\hat{X}_i$ and $\hat{X}_j$ [8].

When the positional fluctuations of $C^{\alpha}$ atoms are of the interest, Eqn. (2.6) can be modified to the form of the correlation of the positional fluctuations of $i^{th}$ and $j^{th}$ residues as follows:

$$\left\langle \Delta R_i\Delta R_j^T\right\rangle = \sum\left(\hat{R}_i-R_i\right)\left(\hat{R}_j-R_j\right)^T f\left(\hat{U},\hat{V},\hat{R}\right) \tag{2.7}$$

where the superscript T denotes the transpose since the position of a $C^{\alpha}$ atom is a vector quantity formed by x, y and z coordinates [9].

The derivative of Eqn. (2.5) with respect to $\left(\frac{F}{T}\right)_i$ leads to:

$$\begin{aligned}\frac{\partial f\left(\hat{U},\hat{V},\hat{R}\right)}{\partial\left(F/T\right)_i} &= k_B^{-1}\left(-\frac{\partial\left(S-\frac{1}{T}U-\frac{P}{T}V+\frac{F}{T}R\right)}{\partial\left(F/T\right)_i}-\hat{R}_i\right)f\left(\hat{U},\hat{V},\hat{R}\right) \\ &= k_B^{-1}\left(\hat{R}_i-R_i\right)f\left(\hat{U},\hat{V},\hat{R}\right)\end{aligned} \tag{2.8}$$

By combining Eqns. (2.8) and (2.7), the following equation is obtained:

$$\begin{aligned}\left\langle \Delta R_i \Delta R_j^T \right\rangle = k_B T \sum\left(\hat{R}_j - R_j\right)\frac{\partial f\left(\hat{U},\hat{V},\hat{R}\right)}{\partial F_i} &= k_B T \frac{\partial}{\partial F_i}\left\langle \hat{R}_j - R_j \right\rangle \\ &\quad - k_B T\left\langle \frac{\partial}{\partial F_i}\left\langle \hat{R}_j - R_j \right\rangle \right\rangle\end{aligned} \tag{2.9}$$

In Eqn. (2.9), the first term of the right hand side vanishes because $\left\langle \hat{R}_j - R_j \right\rangle$ is the first moment of the positional fluctuation with a value of zero [8]. Moreover, $\hat{R}_j$ of the second term is independent of $\mathbf{F}_i$ [9]. Therefore Eqn. (2.9) can be simplified to Eqn. (2.10) [9]:

$$\left\langle \Delta R_i \Delta R_j^T \right\rangle = k_B T\left(\frac{\partial R_j}{\partial F_i}\right)_{T,P,F,i\neq j} \tag{2.10}$$

In general, a recursion relation can be obtained that relates higher order correlation moments to lower order moments [8]:

$$\left\langle \phi \Delta \Phi_k \right\rangle = -k_B \frac{\partial}{\partial \Psi_k} \left\langle \phi \right\rangle - k_B \left\langle \frac{\partial \phi}{\partial \Psi_k} \right\rangle \tag{2.11}$$

where $\phi$ is a higher order product of the fluctuations of the extensive variables $\Delta U, \Delta V, \Delta R$ such as $\Delta R_i \Delta R_j \Delta R_k ...$ $\Phi_k$ represents any of the aforementioned extensive variables and $\Psi_k$ is the corresponding intensive parameters.

Inspired by the basic fluctuation theory explained above, the fluctuation dynamics of the proteins can be investigated practically from the molecular dynamics simulations. An efficient technique to analyze the fluctuation correlations between the residues from the molecular dynamics trajectories is the modal analysis. Modal analysis is a detailed extension of Principal Component Analysis (PCA). However, in modal analysis, the aim is to reveal the local fluctuation dynamics of the proteins at the residue level instead of expression of the global motion by few largest eigenvalues that is adopted in PCA.

The basic theory is the same as PCA with some extensions. The initial step is the construction of the covariance matrix, C, by means of the positional fluctuation matrix ΔR. Subsequently the eigenvalue decomposition (EVD) is carried out to obtain the eigenvectors as the column vectors of V matrix and the eigenvalues that are the diagonal elements of Q matrix. The eigenvalues are the modes and the eigenvectors are the vectors that represent motions of the associated modes.

$$C = \left\langle \Delta R \Delta R^T \right\rangle = VQV^T \tag{2.12}$$

where $\langle\ \rangle$ represents the time average and T stands for the transpose.

C matrix is a 3Nx3N square matrix, where N is the number of atoms. Thus, EVD reveals 3N eigenvalues for an N-particle system, large values representing the slow modes thus associated with the global motion, and small values representing the fast modes associated with the local fluctuations [9].

After obtaining the modes, the fluctuations at the modal space are obtained by the projection, described in the following equation [9].

$$\Delta r = \langle C \rangle^{-1/2} \Delta R = \langle \Delta R \Delta R^T \rangle^{-1/2} \Delta R \tag{2.13}$$

where, Δr is the 3NxM transformed fluctuation matrix with M being the number of the time steps in the trajectory.

By the use of Eqns. (2.13) and (2.14), Δr matrix can be expressed in terms of the eigenvalues and eigenvectors as in Eqn. (2.15) [9].

$$\langle \Delta R \Delta R \rangle^{-1/2} = diag\lambda^{-1/2} V^T \tag{2.14}$$

$$\Delta r = diag\lambda^{-1/2} V^T \Delta R \tag{2.15}$$

where, diagλ are the diagonal elements of the Q matrix, thus the eigenvalues.

In order to study the dynamic characteristics by the correlations at specific modes, back projection to the real space is carried out by performing Eqn. (2.16) [9].

$$\Delta R_k = V_k diag\lambda_k^{1/2} \Delta r \tag{2.16}$$

where, subscript k stands for the $k^{th}$ mode. $\Delta R_k$ matrix is a 3NxM matrix that reflects the fluctuations in the real space originated from the $k^{th}$ mode.

After obtaining the real space fluctuations for the specific modes, the dynamics of the system can be revealed by the correlations given in Eqns. (2.10) and (2.11).

# Chapter 3

## STATISTICAL THERMODYNAMICS OF BINDING I

The correlations between the residue fluctuations and the energy fluctuations have currently become one of the most important topics in statistical thermodynamics since these correlations involve vast amount of information about the functional roles of the residues of a protein in recognition, catalysis and binding.

The energy fluctuations in a protein can originate from the positional fluctuations of the residues or from an external source that transfer energy to the system [10]. Since the slow modes represent the global motions of a protein and the fastest modes indicate the local dynamics, the convenient way of understanding the energetic relations of a protein at the residue level is the investigation of the fastest modes. It has been shown that the residues that are mostly affected by the energy fluctuations, also termed as highly excitable residues, are the ones that are correlated with most of the remaining residues of the protein at the fastest modes [10]. Furthermore, these highly excitable residues were shown to play important roles in binding or to be located at the neighborhood of the binding site [10].

Based on Eqn. (2.11) derived in Chapter 2, the correlation between the energy fluctuations and the positional fluctuations of the residues i and j can be derived as follows [10].

$$\left\langle \Delta U \Delta R_i \Delta R_j^T \right\rangle = \sum \left( \hat{U} - U \right) \left( \hat{R}_i - R_i \right) \times \left( \hat{R}_j - R_j \right)^T f\left( \hat{U}, \hat{V}, \hat{R} \right) \qquad (3.1)$$

$$\left\langle \Delta U \Delta R_i \Delta R_j^T \right\rangle = \left(k_B T\right)^2 \left( \frac{\partial^2 U}{\partial F_i \partial F_j} \right) \tag{3.2}$$

Eqn. (3.3) can be obtained by performing the differentiation given in Eqn. (3.2) [10].

$$\left\langle \Delta U \Delta R_i \Delta R_j^T \right\rangle = k_B T \left\langle \Delta R_i \Delta R_j^T \right\rangle \tag{3.3}$$

The above relation indicates that the fluctuations in energy are distributed to the residues according to the second correlation moment of the positional fluctuations. Thus, Eqn. (3.3) is a way to reveal how the residues of a protein are excited by the energy fluctuations [10].

Another and more relevant way of explaining the relationship between the positional fluctuations and the energy fluctuations at the residue level is the distance fluctuation term which is given with Eqns. (3.4) and (3.5).

$$\left\langle \Delta R_{ij}^2 \right\rangle = \left\langle \left( \Delta R_i - \Delta R_j \right)^2 \right\rangle = \left\langle \Delta R_i^2 \right\rangle - 2 \left\langle \Delta R_i \Delta R_j \right\rangle + \left\langle \Delta R_j^2 \right\rangle \tag{3.4}$$

$$\left\langle \Delta U \left( \Delta R_{ij} \right)^2 \right\rangle = kT \left\langle \left( \Delta R_{ij} \right)^2 \right\rangle \tag{3.5}$$

Eqn. (3.4) shows the expansion of the distance fluctuation between the residues i and j, and Eqn. (3.5) represents that the distance fluctuations are affected by the energy fluctuations of the system directly proportional to the amount of the distance fluctuations [10].

There are three contributions to the distance fluctuation term given in Eqn. (3.4). If the residues i and j exhibit large positional fluctuations, then the distance fluctuation has a greater value originating from the first and the third terms. Moreover, a greater negative $\left\langle \Delta R_i \Delta R_j \right\rangle$ value raises the $\left\langle \Delta R_{ij}^2 \right\rangle$ value [11]. Thus, greater $\left\langle \Delta R_i^2 \right\rangle$ and $\left\langle \Delta R_j^2 \right\rangle$ values and a large negative $\left\langle \Delta R_i \Delta R_j \right\rangle$ determine the influence of the energy fluctuations on the residues i and j.

Another outcome of Eqn. (3.5) is that if a residue i is highly correlated with the remaining residues j through Eqn. (3.5) at the fastest modes than it is related to the binding site of the protein [10]. Thus, the study of the distance fluctuations at the fastest modes gives insight to the local dynamics of a protein system as well as energetics at the residue level.

# Chapter 4

## STATISTICAL THERMODYNAMICS OF BINDING II

Binding of a ligand to a binding site of a receptor protein is a very complex mechanism involving vast amount of molecular interactions. Statistical thermodynamics is the most efficient scope of mathematics and thermodynamics to understand not only the entropic and energetic events taking place but also the molecular contributions to the mechanism.

A protein molecule having 2 identical binding sites can be represented as in Figure 4.1. below.

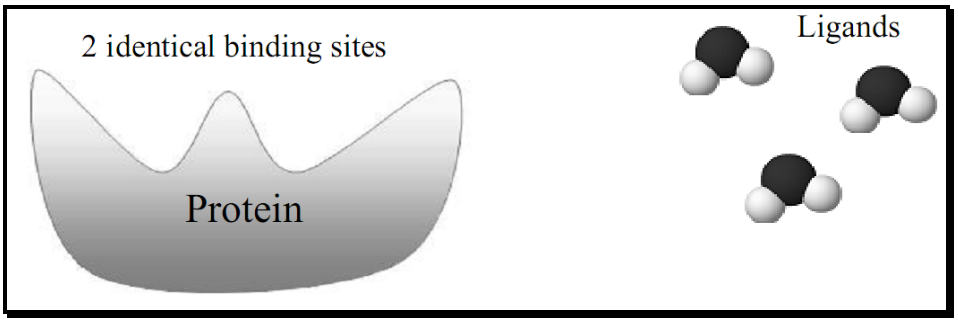


**Figure 4.1.** Protein with 2 binding sites and ligands to bind these sites

For the identical and distinguishable sites, there are four different states of binding. These are; no ligand, 1 ligand bound to the 1$^{st}$ site and no ligand on the 2$^{nd}$ site, 1 ligand

bound to the 2nd site and no ligand on the 1st site, and 2 ligands bound to each site. These states can be represented graphically as shown in Figure 4.2. below.

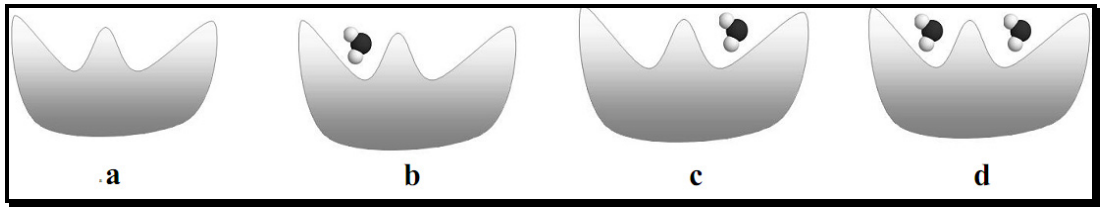


**Figure 4.2. a)** protein with no ligand **b)** one ligand bound to the 1st site **c)** one ligand bound to the 2nd site **d)** both sites are occupied

If the system above is assumed to be a grand canonical ensemble that is exposed to the temperature and particle reservoirs, the partition function, that describes the statistical characteristics of the system in thermodynamic equilibrium can be written as [12]:

$$\xi(T,\mu) = Q(0,0) + \left[Q(1,0) + Q(0,1)\right]\lambda + Q(1,1)\lambda^2 \qquad (4.1)$$

where Q(0,0), Q(1,0), Q(0,1) and Q(1,1) are the partition functions for each binding state shown in Figure 4.2., 0 representing the empty site and 1 representing the occupied site, respectively. These partition functions include the energy levels of each binding site and the adsorption energies of each ligand. λ is called the absolute activity and depends on the chemical potential of the ligands and the temperature of the system [12].

After obtaining the partition function for the system, derivations of the statistical relations are straightforward. Distribution functions which in fact are the probabilities of each binding state can be derived by the use of Eqn. (4.1) [12].

The probability of finding no site occupied is given by [12]:

$$P(0,0) = \frac{1}{\xi} \tag{4.2}$$

The probability of finding the 1st site occupied and the 2nd site empty is [12]:

$$P(1,0) = \frac{Q(1,0)\lambda}{\xi} \tag{4.3}$$

Similar to previous one, the probability of finding the 2nd site occupied and the 1st site empty is given as [12]:

$$P(0,1) = \frac{Q(0,1)\lambda}{\xi} \tag{4.4}$$

And lastly the probability of finding both sites occupied by a ligand which is also termed as joint probability, can be obtained as [12]:

$$P(1,1) = \frac{Q(1,1)\lambda^2}{\xi} \tag{4.5}$$

Conditional probability is an important probabilistic term that is practical for the derivation of the correlation function. The conditional probability for the above system, that is the probability of binding of a ligand to the 2nd site given that the 1st site is already occupied, can be derived by the use of Eqns. (4.3) through (4.5) and is given in Eqn. (4.6) [12].

$$P(1|1) = \frac{P(1,1)}{P(1)} \tag{4.6}$$

where P(1) is the probability of finding 1 site occupied irrespective of the type of the site and can be obtained by the summation of Eqns. (4.3) and (4.4).

The correlation function at infinite dilution, $g^o(1,1)$, is a measure of the deviation of the conditional probability from the unconditional probability and can be obtained as Eqn. (4.7) below [12].

$$g^o(1,1) = \frac{P(1|1)}{P(1)} = \frac{P(1,1)}{P(1)^2} \tag{4.7}$$

In the above equation the right hand side is obtained by the substitution of Eqn. (4.6) for the conditional probability term.

The correlation function has a determinant role for the cooperativity between the two ligands. If the value of $g^o(1,1)$ is 1, then the two ligands don't interact with each other. That is the binding of a ligand to one site isn't affected by the occupancy of the other binding site. However, when there is cooperativity between the ligand molecules, the conditional probability deviates from the unconditional probability. Hence, $g^o(1,1)$ having a value greater than 1 represents the positive cooperativity between the ligands, whereas a value smaller than 1 indicates the negative cooperativity [12]. In detail, the positive cooperativity infers that the binding of a ligand to a site attracts the other ligands for the other binding sites, whereas the negative cooperativity has a reciprocal meaning.

By studying the correlation function, the cooperative manner between the modifications of the lysine residues can be conceived at the statistical thermodynamics basis.

# Chapter 5

## MODEL AND COMPUTATIONS

Three types of 40-residue histone H3 tails were selected to carry out molecular dynamics simulations based on the cross-regulation knowledge explained in Chapter 1. These were the unmodified histone H3 tail having no chemical modification on the residues, the tri-methylated lysine 4 (H3K4me3) tail having 3 methyl groups added to N-ε atom of the $4^{th}$ lysine residue and the tri-methylated lysine 9 (H3K9me3) tail having 3 methyl groups similarly added to the $9^{th}$ lysine residue. The 40-residue part of the N-terminal of histone H3 that protrudes from the nucleosome surface was obtained from the crystal structure of the nucleosome with the PDB ID 1KX5 [13]. The addition of methyl groups to the lysine side-chains was performed by Accelrys Discovery Studio Visualizer program [14]. Since these chemical modifications divert the partial charges and the chemical parameters of the lysine residues, parametrization procedure was carried out for the methylated lysine residues as explained in detail in the next section. For each of the three H3 tails, 6 MD simulations were carried out, starting from different initial configurations to account for agreeable sampling. These different configurations were obtained by taking random snapshots from a 2 ns simulated annealing trajectory (data not shown).

## 5.1 Parametrization of Methylated Lysine Residues

As mentioned above, addition of methyl groups to the lysine residues causes deviations of the chemical properties and especially the partial charges from the unmodified lysine residues. Thus, parametrization procedure was carried out in order to obtain simulation parameters of the methylated lysine residues.

The chemical parameters for the tri-methylated lysine residues such as bond lengths, bond angles, dihedral angles were obtained from the Duan *et. al.* force field (2003) [15]. For the partial charges, Ante_RED, RED III and Gaussian03 programs were used to derive RESP charges [16, 17].

RESP charge derivation procedure was performed in three steps; geometry optimization, Molecular Electrostatic Potential (MEP) computation using the optimized geometry obtained in the first step and fitting the charges centered on the atoms to the MEP calculated in the second step [16].

As the initial step, Ante_RED program was executed to obtain input files for the geometry optimization and MEP computation [16]. Subsequently, the geometry optimization was carried out via Gaussian03. Furthermore, RED-III program was used for MEP computation of the modified lysine residues with the optimized geometries [16]. MEP computation was performed by taking the dielectric constant (ε) as 4 under continuum solvent conditions and with B3LYP/cc-pVTZ basis set that was used in Duan *et. al.* force field (2003) [15]. As the final step RESP charge fitting was carried out and the partial charges for each atom in the modified lysine residues was obtained for MD simulations.

## 5.2 Molecular Dynamics Simulations

Molecular dynamics (MD) simulations are used to compute the time dependent behavior of biological molecules. They provide detailed information about the fluctuations and conformational changes of the proteins and nucleic acids hence being very useful tool to investigate the structure, function, dynamics and thermodynamics of the biological molecules. MD algorithm is based on the Newton's second law of motion. For an N-particle system, all the interactions are described by this equation of motion as follows.

$$F_i = m_i \cdot a_i = m_i \cdot \frac{dv_i}{dt} = m_i \cdot \frac{d^2 x_i}{dt^2} \qquad i = 1, 2, ..., N \tag{5.1}$$

where $m_i$ and $x_i$ are the mass and position of $i^{th}$ atom respectively and $F_i$ is the force acting on that atom. The force $F_i$ is obtained as the derivative of the potential function $\Phi(x_1, x_2, \ldots, x_N)$ as given in Eqn. (5.2).

$$F_i = -\frac{\partial \Phi}{\partial x_i} \tag{5.2}$$

The potential function for the system accounts for all the particle interactions of the system and is given as:

$$\Phi_{total} = \Phi_{non-bonded} + \Phi_{bonded} + \Phi_{other} \tag{5.3}$$

Based on the theory, six MD simulations were carried out for each of the three H3 tails. The initial configurations of each of 6 simulations for the tails were the same e.g. the $1^{st}$

simulations of each tail were started with the same initial configuration. The simulations were performed in explicit solvent with NAMD 2.5 package by using Duan *et. al.* force field (2003) [15,18]. In order to be consistent with the cellular conditions, isothermal-isobaric (NTP) ensemble was used for the simulations with the periodic boundary conditions. The temperature and pressure were set to 310 K and 1 bar, respectively. Computations of non-bonded interactions were carried out with a cut-off distance of 12 Å. The Particle Ewald sum was used for the calculation of long-range forces, thus minimizing the error due to the cut-off distance in the periodic systems. For each simulation 1000 step energy minimization was performed. The simulations were recorded for 7 ns with an integration time step of 2 fs.

### 5.3 Modal Analysis of MD Trajectories

The modal analysis of the MD trajectories was performed for the investigation of the local fluctuation dynamics of the H3 tails. For each simulation, the first 2 ns part was taken as the time for the system to reach equilibrium and the last 5 ns part of 7 ns simulations was extracted for the modal analysis. The fluctuation matrices, $\Delta R$, were obtained for $C^{\alpha}$ coordinates of 40 residues. Subsequently, the covariance matrices were computed by using Eqn. (2.12) given in Chapter 2 and eigenvalue decomposition (EVD) was applied to the covariance matrices obtained. The first six eigenvalues of each simulation were either zero or close to zero due to the removal of overall translation and rotation of the simulations. The fluctuations at modal space were procured by applying Eqn. (2.14). For the analysis of the local fluctuations at the real space, back projection was accomplished for the fastest 18 modes by the use of Eqn. (2.15). The reason for adopting the fastest 18 modes was the linear attitude of the eigenvalues for these indexes as well as the significant revelation of the modification sites on H3 tails by these modes.

Subsequent to the back projection, $\left\langle \Delta R_{ij}^2 \right\rangle$ correlation maps were obtained by the help of Eqns. (2.16) and (3.4) conferred in Chapter 2 and Chapter 3. As six simulations were carried out for each tail, the averages were taken to obtain single result for each tail.

### 5.4 Correlation Function and Cooperativity

As $\left\langle \Delta R_{ij}^2 \right\rangle$ correlations are the indications for the distribution of energy fluctuations to the system, they were treated as the correspondents of the partition function. Based on this assumption, normalized $\left\langle \Delta R_{ij}^2 \right\rangle$ correlations were taken to be equivalent to the joint probabilities described in Chapter 3. Broadly, for further analysis, 2 identical and distinguishable binding sites were assumed to be a residue pair, meanwhile the binding of a ligand to a site was taken to be a modification of a residue, thus a site occupied by a ligand being a modified residue.

Further correlation function analysis were focused on the 3 residue pairs K4-K9, K4-K14 and K9-K14, each residue of which were proven to prevent/promote each others modifications as stated in Chapter 1.

Starting from the conditional probabilities indicated in Eqn. (3.6), the correlation function was computed by the use of Eqn. (3.7) for each pair mentioned above. As the initial step, the conditional probability given in Eqn. (3.6) was modified as shown in Eqn. (5.4).

$$P(i \mid j) = \frac{P(i,j)}{P(j)} \tag{5.4}$$

where i and j were the residues of the aforementioned residue pairs, P(i,j) was taken to be the normalized $\left\langle \Delta R_{ij}^2 \right\rangle$ correlations, indicating the probability of interaction between the residues i and j, and P(j) was the sum of the normalized $\left\langle \Delta R_{ij}^2 \right\rangle$ correlations of the residue j, which was taken to be the methylated residue. For the correlation functions of K4-K9 and K4-K14 pairs, H3K4me3 tail was used whereas for K9-K4 and K9-K14 pairs H3K9me3 tail was used in computations.

The correlation function given in Eqn. (3.7) was also modified to the Eqn. (5.5) below.

$$g^o(i,j) = \frac{P(i \mid j)}{P(j)} \tag{5.5}$$

where P(j) similarly was taken to be the sum of the normalized $\left\langle \Delta R_{ij}^2 \right\rangle$ correlations for the K4 residue in the case of H3K4me3 and the K9 residue in the case of H3K9me3.

Based on the results obtained for the correlation function $g^o(i,j)$, the cooperativity between the modifications of the residue pairs was investigated.

# Chapter 6

# RESULTS

Three types of histone H3 tails were prepared for MD simulations. Those were the unmodified H3 tail, the H3K4me3 tail and the H3K9me3 tail. The choice for the simulation tails was originated form the cooperativity knowledge between the K4, K9 and K14 modifications. Figures 6.1. through 6.3. show the structures of the three H3 tails in ribbon representation, where the three residues of interest (K4, K9, K14) represented with ball and stick form.

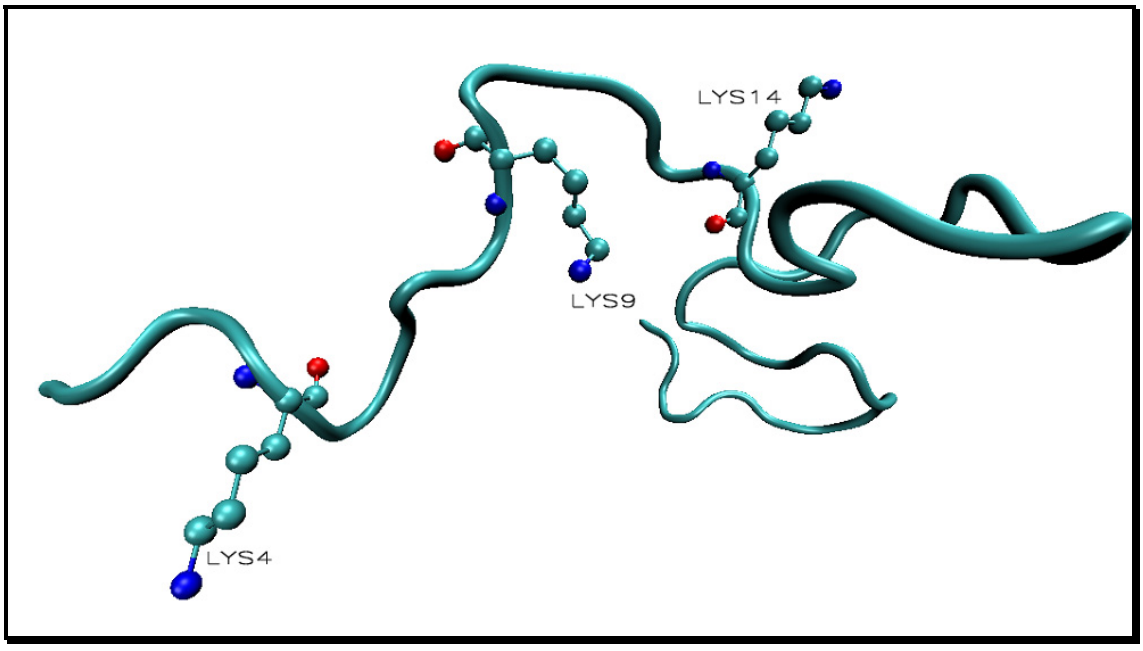


**Figure 6.1.** Structure of the unmodified H3 tail

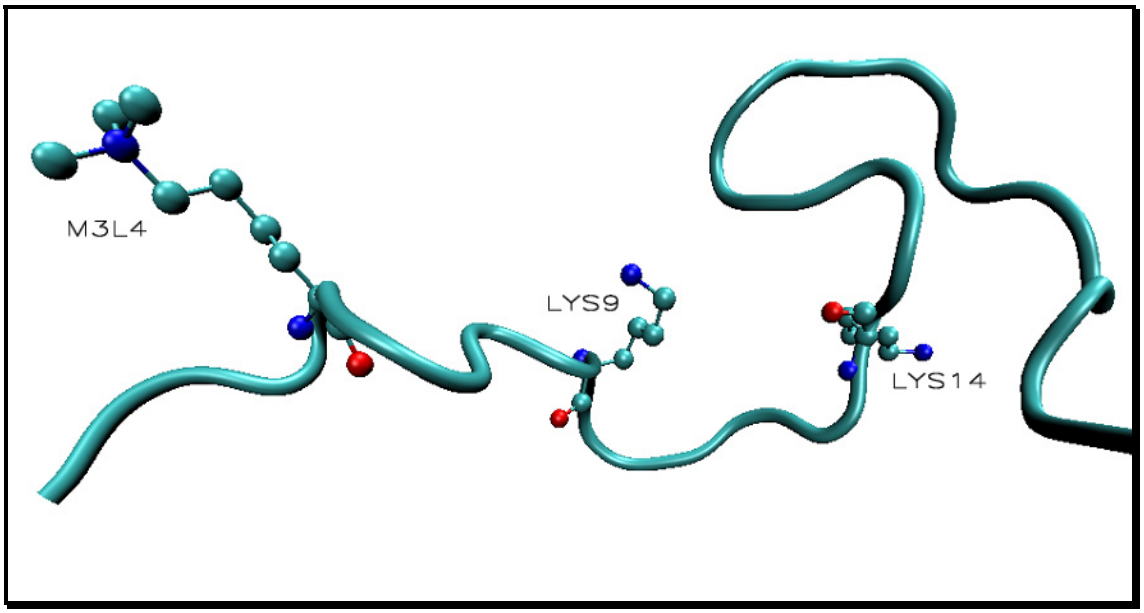


**Figure 6.2.** Structure of the H3K4me3 tail with three methyl groups on K4

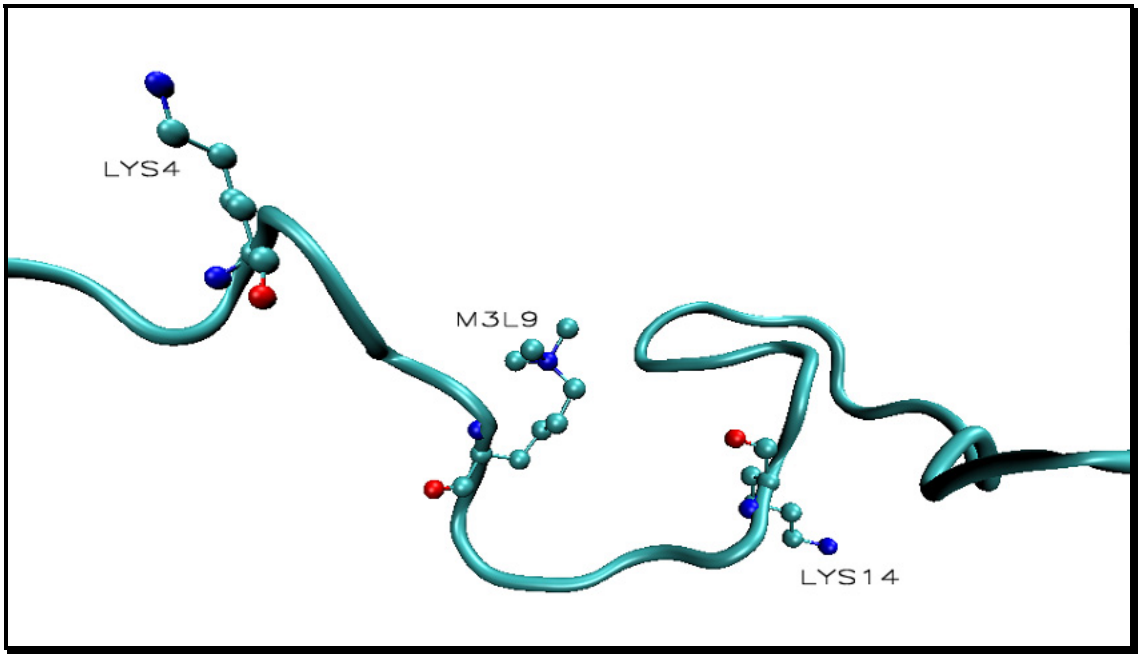


**Figure 6.3.** Structure of the H3K9me3 tail with three methyl groups on K9

After the 7 nanosecond MD runs, modal analysis was performed to the last 5 nanosecond part of each trajectory. In the modal analysis, the fastest 18 modes were used to investigate the local dynamics of the H3 tails. The fastest 18 modes were adopted due to the linear behavior of these eigenvalues in the logaritmic plot of the whole eigenvalues. Following the linearity, a jump can be observed from each plot, showing the transition to the global motion. Figures 6.4. to 6.6. display these logaritmic plots of the eigenvalues starting from the $7^{th}$ eigenvalue, as the first 6 eigenvalues have zero value due to the removal of overall rotation and translation of the H3 tails.

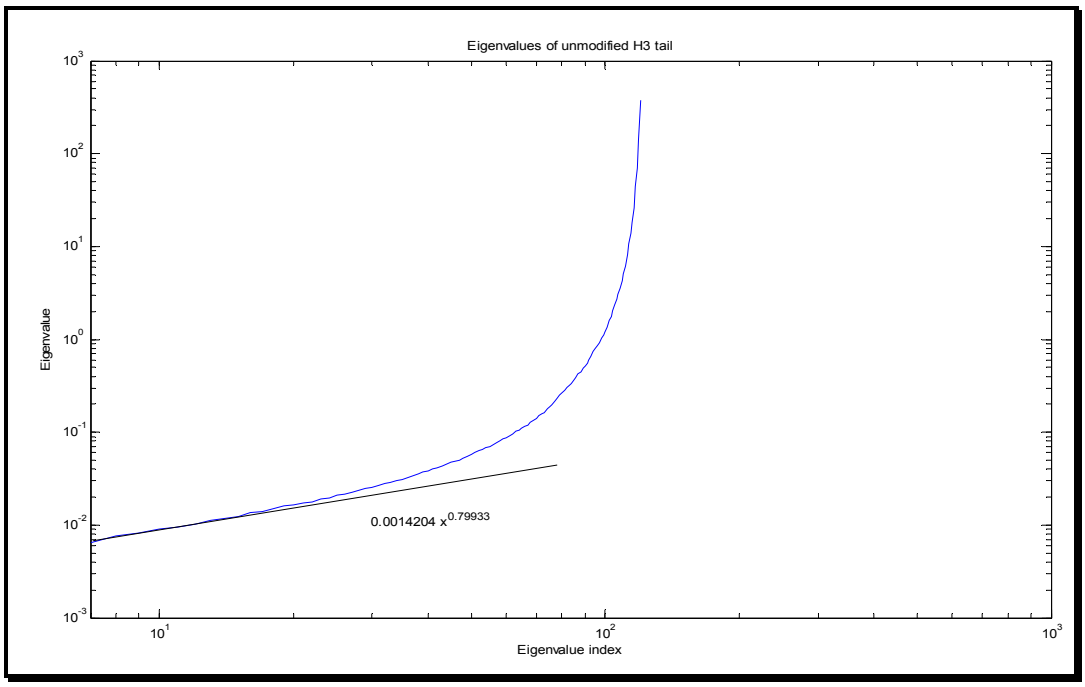


**Figure 6.4.** Logaritmic plot of eigenvalues for unmodified H3 tail

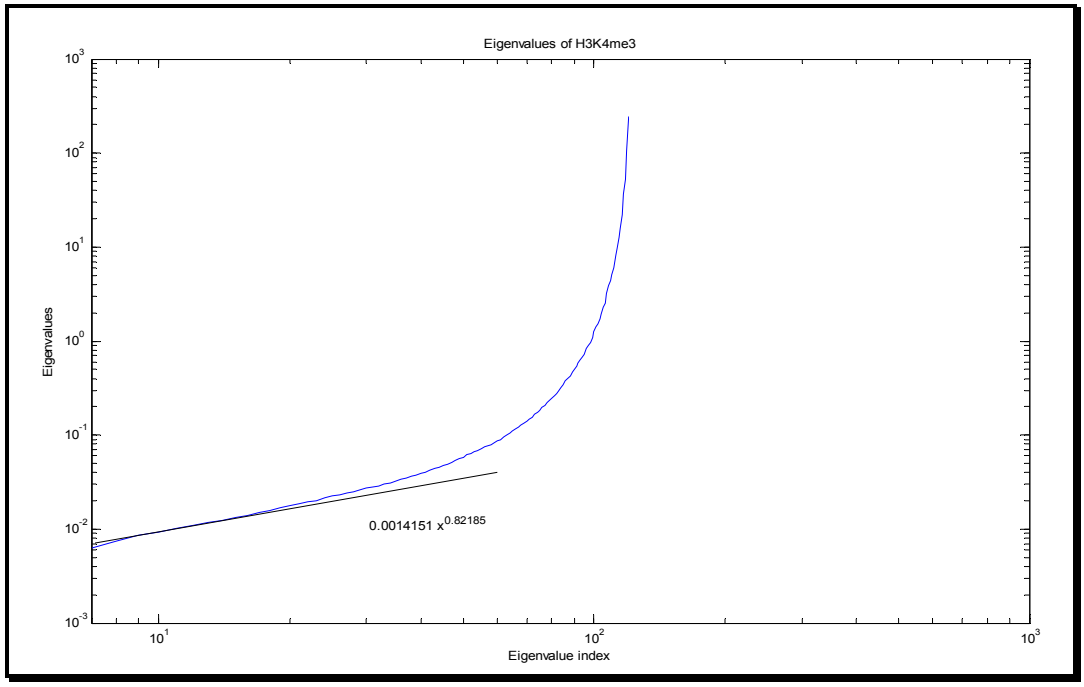


**Figure 6.5.** Logaritmic plot of eigenvalues for H3K4me3 tail

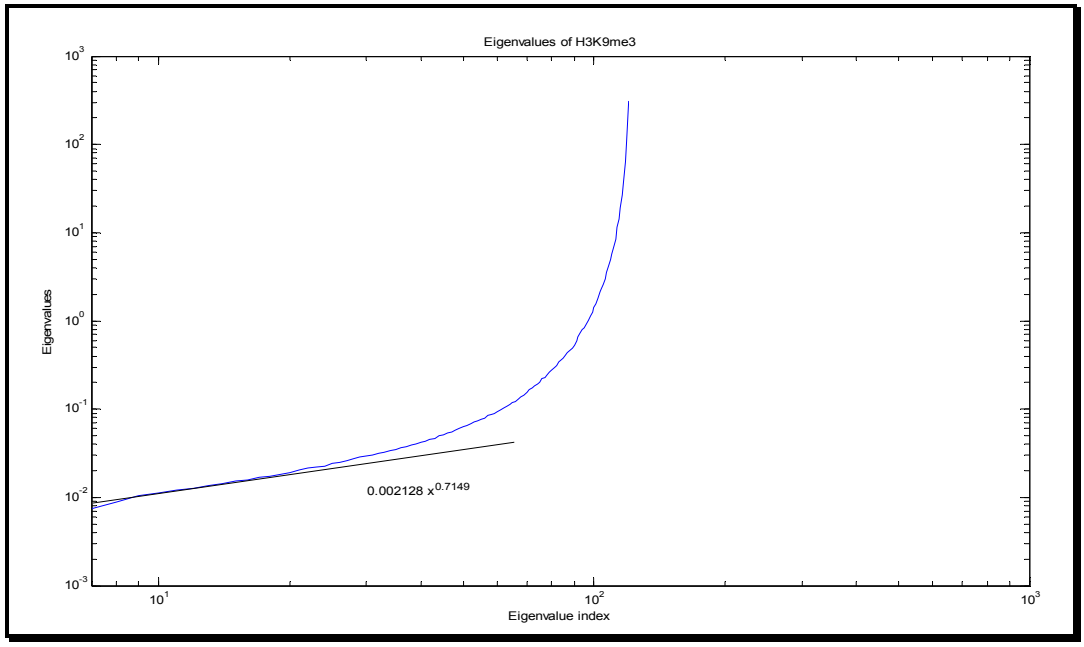


**Figure 6.6.** Logaritmic plot of eigenvalues for H3K9me3 tail

As described in Section 5.3, the modal analysis for the 3 types of H3 tails was accomplished and $\langle \Delta R_{ij}^2 \rangle$ correlation maps were obtained. In Chapter 3 it was mentioned that, $\langle \Delta R_{ij}^2 \rangle$ correlations mark either the binding site residues or their neighboring residues. These indications appear as vertical contours in the correlation maps as the residues of interest correlate with the remaining residues. Figures 6.7. through 6.9. display $\langle \Delta R_{ij}^2 \rangle$ residue correlation maps obtained for the three H3 tails.

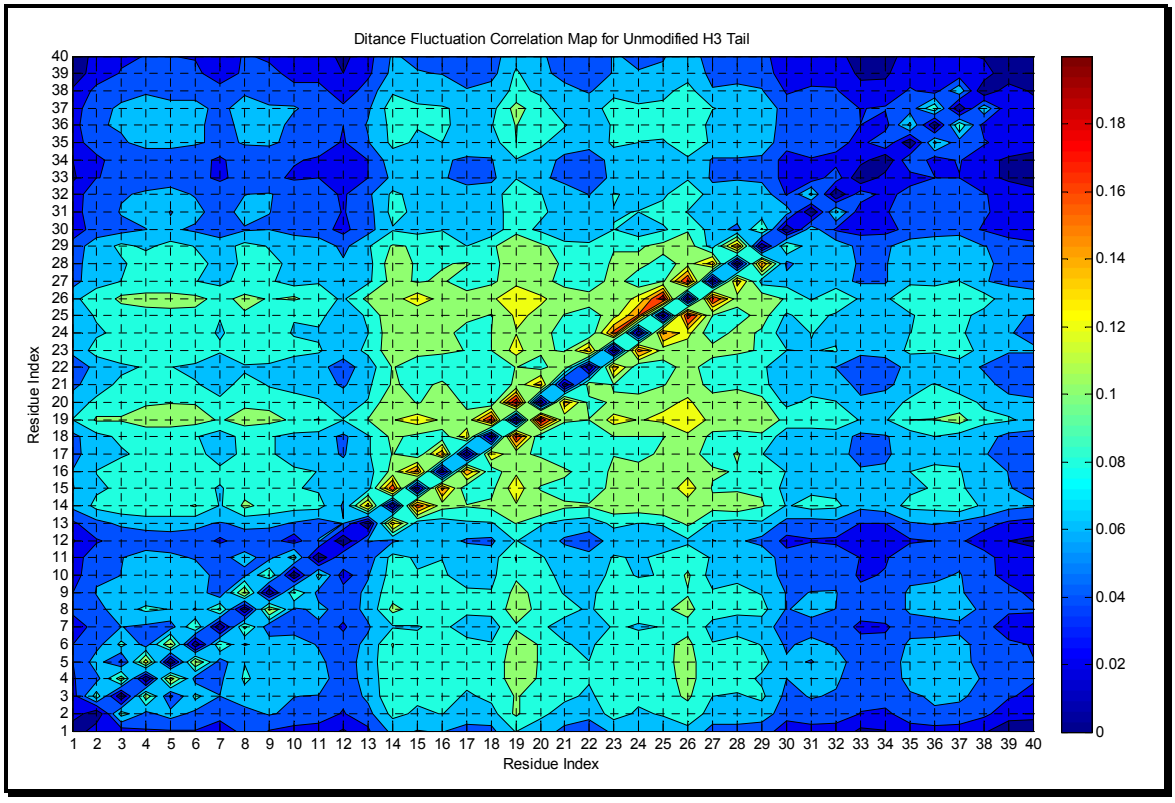


**Figure 6.7.** $\langle \Delta R_{ij}^2 \rangle$ Residue correlation map for the unmodified H3 tail

In Figure 6.7. above, the $\left\langle \Delta R_{ij}^2 \right\rangle$ correlation map for the unmodified H3 tail reveals the residues that are correlated with almost all of the tail residues. These are the residues at the N-terminal region K4 to T6, R8 & K9, K14 to A29, and K36 & K37. These strongly correlated residues are either a modification site on H3 tail or the neighboring residues of a modification site. The most significant correlations are observed for the Q19 and R26 which suggests that these residues are the most fluctuating ones through the whole tail. Q19 is the neighbor residue of K18 which is an acetylation site and R26 is the dimethylation site as well as being a neighbor to another methylation site, K27. Besides, all the three target residues (K4, K9 and K14) are observed to be correlated with most of the tail residues thus, being the modification sites. The correlation contours of K4 through T6 verify K4 as a binding site, whereas the correlations through the residues R8 and K9 signify these modification sites. Furthermore, significant contours around K14 also points out that residue as a modification site.

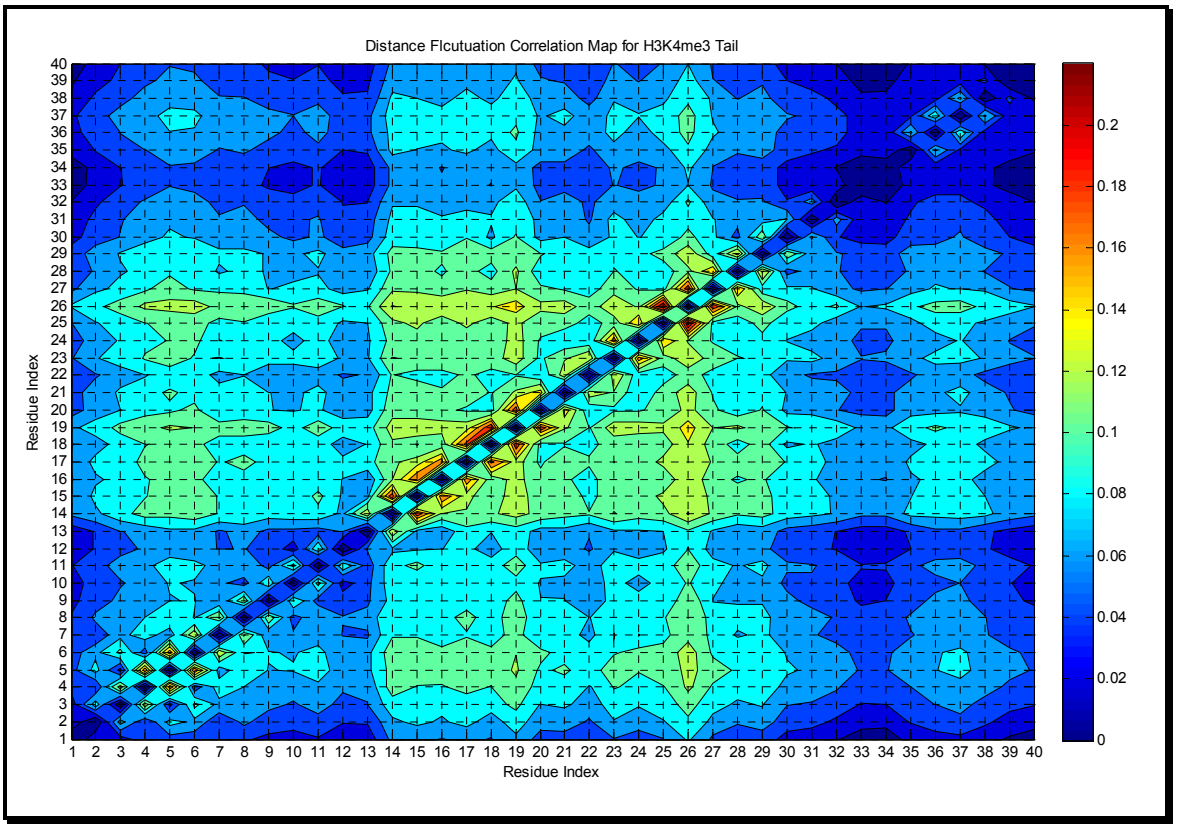


**Figure 6.8.** $\left\langle \Delta R_{ij}^2 \right\rangle$ Residue correlation map for the K4me3 H3 tail

The $\left\langle \Delta R_{ij}^2 \right\rangle$ correlation map for the H3K4me3 tail reveals similar results to the unmodified H3 tail with some significant exceptions. The residues K4 to T6, K14 to A29 and K37 have correlations with the whole tail residues as it was observed in the unmodified case. However, it can clearly be recognized that the correlations for the residue K9 were decreased in the case of tri-methylation of K4. This decrease gives a hint about the communication between K9 and K4 in the H3K4me3 case. Moreover, it can also be observed that the tri-methylation increases the correlations of the residues K4 to T6, which

proposes that the addition of three methyl groups disrupts the stability of K4 and the adjacencies, providing them a more dynamic structure.

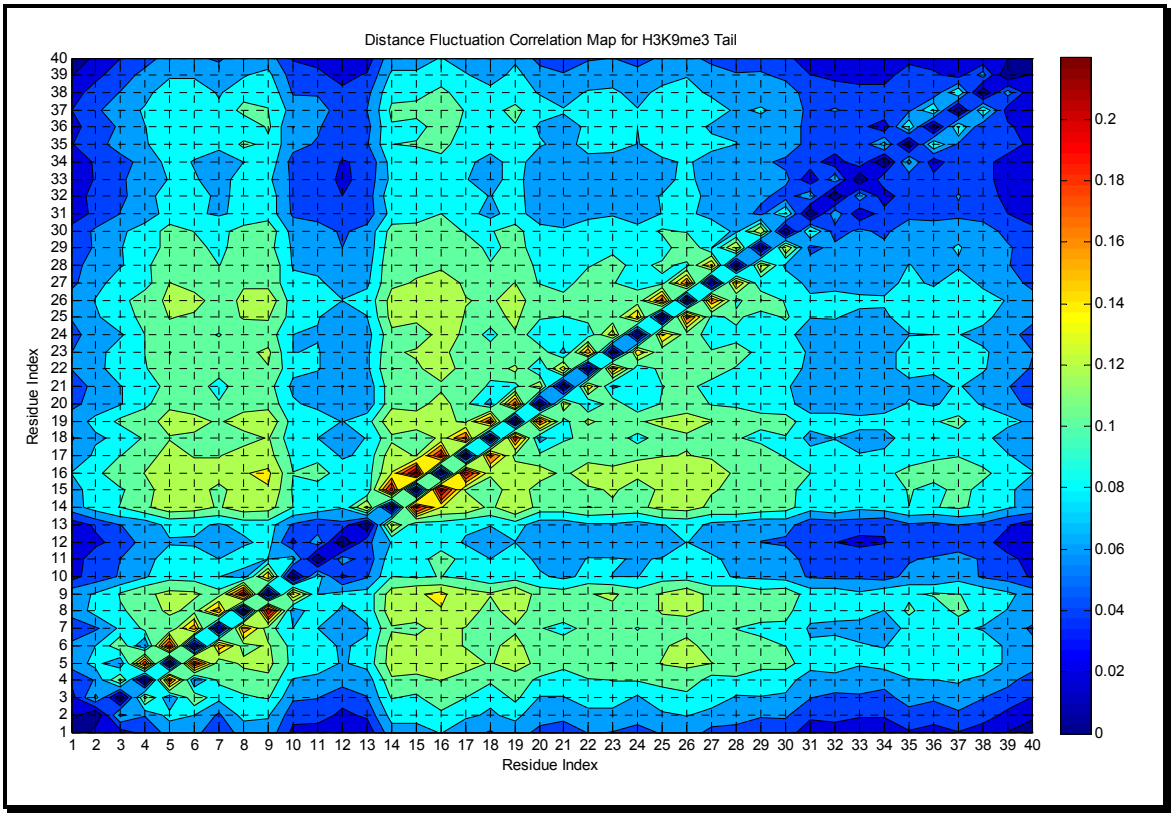


**Figure 6.9.** $\left\langle \Delta R_{ij}^2 \right\rangle$ Residue correlation map for the K9me3 H3 tail

The $\left\langle \Delta R_{ij}^2 \right\rangle$ correlation map for the H3K9me3 tail brings out that the addition of methyl groups causes a drastic increase in the correlations of the residues K4 to S10 as well as the remaining residues that were shown to be correlated with the whole tail in the previous cases. However the jump in K4 correlations occurs to be relatively insignificant compared to the rise of the correlations of K9 and other residues in the H3K9me3 case. Thus, based on this comparison, it may still be concluded that the modification of K9 by three methyl

groups interrupts the modification site characteristics of K4. A similar situation is observed for K14. Although the correlations of K14 are monitored to increase in the H3K9me3 case compared to the unmodified and K4me3 cases, the rise is small according to the jump of the K9 correlations, thus pointing out a similar communication between these residues. In Table 6.1. below, the alteration of the correlation values for these three residues can obviously be observed for the H3K4me3, H3K9me3 and unmodified H3 tail.

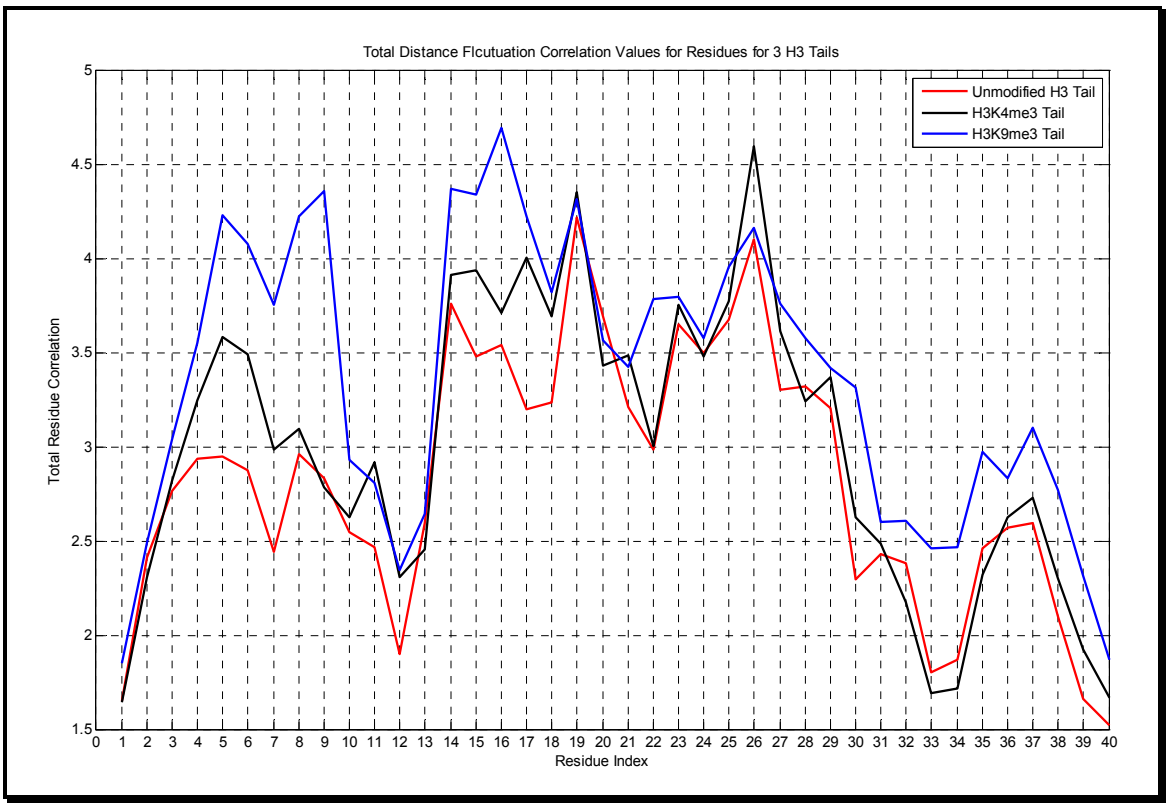


**Figure 6.10.** Sum of $\langle \Delta R_{ij}^2 \rangle$ correlations of residues for three H3 tails

The above figure compares the total correlations of each residue for the unmodified, H3K4me3 and H3K9me3 cases. Broadly, it can be concluded that the methylations on the

H3 tail increase the correlations for the whole tail, especially in the H3K9me3 case, suggesting that these modifications improve the whole dynamics of the H3 tail. For the unmodified tail (red line), the peaks are observed in the K4-Q5, R8-K9, K14, Q19, R26 and K36-K37 regions. The minimum correlations for all three tails are observed for the residues G12, G33 and G34 which can be explained with glycine being the smallest amino acid, thus having the greatest stability due to the lack of a dynamic side chain. Furthermore, G12-G13 and G33-G34 regions of the H3 tail are reasonably expected to be the most stable regions as the double occurrence of glycine prevents large fluctuations. Additionally, the magnitudes of K4 and K9 correlations are observed to be close to each other in the unmodified H3 tail case.

The tri-methylation of K4 increased the correlations of residues K4 to R8, T11, G12, K14 to Q19, A21, K23, A25 to K27, A29 to S31, and K36 to R40, compared to the unmodified tail. However, there is a slight decrease in the correlation of K9 which is a potential reason for the inhibition of K9 methylation in the case of K4 methylation.

In the case of the tri-methylated K9, the modification seems to cause a severe increase in the correlations of almost the whole residues. This rise is observed especially for the first 10 residues of the N-terminal, K14 to K18, S22 to A25, and K27 to R40. As expected, the most significant effects of modification are monitored for the correlations of K9 and the neighboring residues.

A more relevant observation is the magnitude of the rise in correlations of the H3K4me3 and H3K9me3 cases in reference to the unmodified tail case. Table 6.1. below summarizes the correlation values and their alterations for the three residues of interest, K4, K9 and K14.

**Table 6.1.** $\left\langle \Delta R_{ij}^2 \right\rangle$ Correlations of residues K4, K9 and K14

| Residue | Unmodified Tail | K4me3 Tail | K9me3 Tail | K4me3-Unmodified | K9me3-Unmodified |
|---|---|---|---|---|---|
| **K4** | 2.9361 | 3.2489 | 3.5537 | 0.3128 | 0.6176 |
| **K9** | 2.8305 | 2.7846 | 4.3559 | -0.0459 | 1.5254 |
| **K14** | 3.7604 | 3.9127 | 4.3702 | 0.1523 | 0.6098 |

According to Table 6.1., it is obvious that the addition of methyl groups either to K4 or K9 mostly raised the $\left\langle \Delta R_{ij}^2 \right\rangle$ correlations of the three residues of interest. The modification of K4 gained an increase in the correlations of K4 and K14, K4 having the greatest rise. However, this modification caused a decrease in the correlation value of K9 compared to the unmodified H3 tail. Moreover, the addition of methyl groups to K9 caused all the three residues to have greater correlation value as well as the most of the tail residues. But the increase is insignificant for K4 and K14 in reference to the rise in K9, which can be related to the cooperativity.

All together these results demonstrate that, by studying the fluctuation dynamics at the fastest modes, the modification sites on the H3 tail can be regionally detected for the unmodified, H3K4me3 and H3K9me3 cases. Moreover, the addition of methyl groups to the lysine residues (K4, K9) interrupts the dynamics of the whole tail residues and the modification site characteristics of the three residues (K4, K9, K14) of the interest.

Starting from the theory interpreted in Chapter 4 and using the hypothetical derivations explained in Section 5.4., the correlation functions for the residue pairs of interest were computed for the H3K4me3 and H3K9me3 cases. Table 6.2. below summarizes the results of the computations.

**Table 6.2.** $g^o(i,j)$ Values computed for the K4me3 & K9me3 tails

| | K4me3 | | | | K9me3 | | | |
|---|---|---|---|---|---|---|---|---|
| **Pairs** | $P(j)$ | $P(i,j)$ | $P\langle i \mid j \rangle$ | $g^o(i,j)$ | $P(j)$ | $P(i,j)$ | $P\langle i \mid j \rangle$ | $g^o(i,j)$ |
| **(K4, K9)** | 0.0271 | 6.2288e-004 | 0.0230 | **0.8478** | 0.0324 | 8.6502e-004 | 0.0267 | **0.8232** |
| **(K4, K14)** | 0.0271 | 9.1509e-004 | 0.0338 | **1.2456** | - | - | - | - |
| **(K9, K14)** | - | - | - | - | 0.0324 | 9.5725e-004 | 0.0295 | **0.9109** |

It was mentioned in Chapter 4 that $g^o(i,j)$ values greater than 1 indicate a positive cooperativity, values smaller than 1 point out a negative cooperativity, whereas values equal to 1 mean no cooperativity. It can clearly be recognized from Table 6.2. that the residue pairs (K4,K9), (K9,K4) and (K9,K14) have the negative cooperativity with the $g^o(i,j)$ values smaller than 1, whereas the residue pair (K4,K14) has the positive cooperativity with a 1.2456 $g^o(i,j)$ value. These results suggest that the cross-talk between the modifications of the aforementioned residues can explicitly be explained by the correlation function, $g^o(i,j)$. Based on this proposal, the promotion between the modifications can be represented with a positive cooperativity, thus a $g^o(i,j)$ value greater than 1, whereas the prevention between the modifications can be indicated with a negative cooperativity between the residues, having a $g^o(i,j)$ value smaller than 1.

Based on these inferences, the cooperativity of the tri-methylated residues with the remaining tail residues, either a binding site or not, were investigated. Table 6.3. and Figure 6.11. summarize the whole tail cooperativity computations for K4 of the H3K4me3 tail, whereas Table 6.4. and Figures 6.12. display the results for K9 of the H3K9me3 tail.

**Table 6.3.** Cooperativity results for K4me3 with rest of the tail residues

| **Residue** | $g^o(i,4)$ | **Cooperativity** | **Residue** | $g^o(i,4)$ | **Cooperativity** |
|---|---|---|---|---|---|
| A1 | 0.5684 | inhibition | S22 | 0.8935 | inhibition |
| R2 | 0.4651 | inhibition | ***K23*** | ***1.1533*** | ***activation*** |
| T3 | 1.5016 | activation | A24 | 1.0719 | activation |
| Q5 | 1.9817 | activation | A25 | 1.1272 | activation |
| T6 | 0.5393 | inhibition | R26 | 1.4121 | activation |
| A7 | 1.0251 | activation | K27 | 1.0974 | activation |
| R8 | 0.9089 | inhibition | S28 | 0.9906 | inhibition |
| S10 | 0.8037 | inhibition | A29 | 1.0475 | activation |
| T11 | 0.8748 | inhibition | P30 | 0.8371 | inhibition |
| G12 | 0.7183 | inhibition | S31 | 0.7797 | inhibition |
| G13 | 0.7405 | inhibition | T32 | 0.7023 | inhibition |
| A15 | 1.1228 | activation | G33 | 0.5682 | inhibition |
| P16 | 1.1207 | activation | G34 | 0.5980 | inhibition |
| R17 | 1.1887 | activation | V35 | 0.6795 | inhibition |
| ***K18*** | ***1.1303*** | ***activation*** | K36 | 0.9063 | inhibition |
| Q19 | 1.2904 | activation | K37 | 0.8478 | inhibition |
| L20 | 1.0570 | activation | P38 | 0.7781 | inhibition |
| A21 | 1.0641 | activation | H39 | 0.6398 | inhibition |
| | | | R40 | 0.5684 | inhibition |

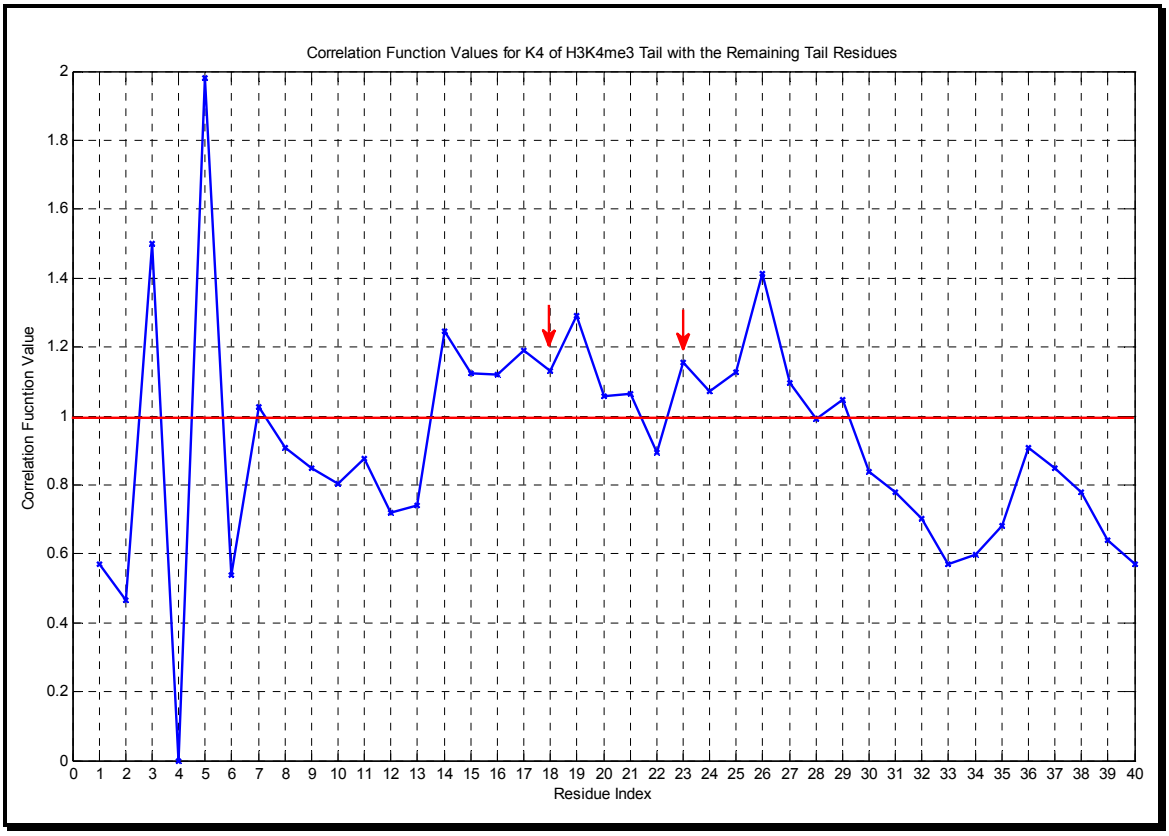


**Figure 6.11.** $g^{o}(i,4)$ Values for K4 of the H3K4me3 tail

It was shown in experimental studies that the tri-methylation of K4 promotes the acetylation through the H3 tail which suggests the activation of all of the acetylation sites that are K9, K14, K18 and K23 [7]. It is inaccurate to examine the acetylation of K9 due to its dual role as a methylation site as well as an acetylation site. However, the remaining residues that are K14, K18 and K23 are only the acetylation sites having no other modification. Previosuly the acetylation of K14 was shown to be promoted by this modification. However, from Table 6.3. and Figures 6.11. above, it is obvious that the other two acetylation sites on the H3 tail, which are K18 and K23, are activated by the methylation of K4 as well.

**Table 6.4.** Cooperativity results for K9me3 with rest of the tail residues

| **Residue** | $g^o(i,9)$ | **Cooperativity** | **Residue** | $g^o(i,9)$ | **Cooperativity** |
|---|---|---|---|---|---|
| A1 | 0.5092 | inhibition | S22 | 0.8446 | inhibition |
| R2 | 0.5819 | inhibition | ***K23*** | ***0.8707*** | ***inhibition*** |
| T3 | 0.7205 | inhibition | A24 | 0.7811 | inhibition |
| Q5 | 0.8887 | inhibition | A25 | 0.8631 | inhibition |
| T6 | 0.8746 | inhibition | R26 | 0.9530 | inhibition |
| A7 | 0.7581 | inhibition | K27 | 0.7960 | inhibition |
| R8 | 1.5574 | activation | S28 | 0.8535 | inhibition |
| S10 | 1.1443 | activation | A29 | 0.7571 | inhibition |
| T11 | 0.5626 | inhibition | P30 | 0.7977 | inhibition |
| G12 | 0.6116 | inhibition | S31 | 0.6369 | inhibition |
| G13 | 0.6286 | inhibition | T32 | 0.6689 | inhibition |
| A15 | 0.9776 | inhibition | G33 | 0.5963 | inhibition |
| P16 | 1.0208 | activation | G34 | 0.6585 | inhibition |
| R17 | 0.9054 | inhibition | V35 | 0.6742 | inhibition |
| ***K18*** | ***0.8613*** | ***inhibition*** | K36 | 0.7295 | inhibition |
| Q19 | 0.9728 | inhibition | K37 | 0.7111 | inhibition |
| L20 | 0.7891 | inhibition | P38 | 0.6816 | inhibition |
| A21 | 0.7698 | inhibition | H39 | 0.6038 | inhibition |
| | | | R40 | 0.5022 | inhibition |

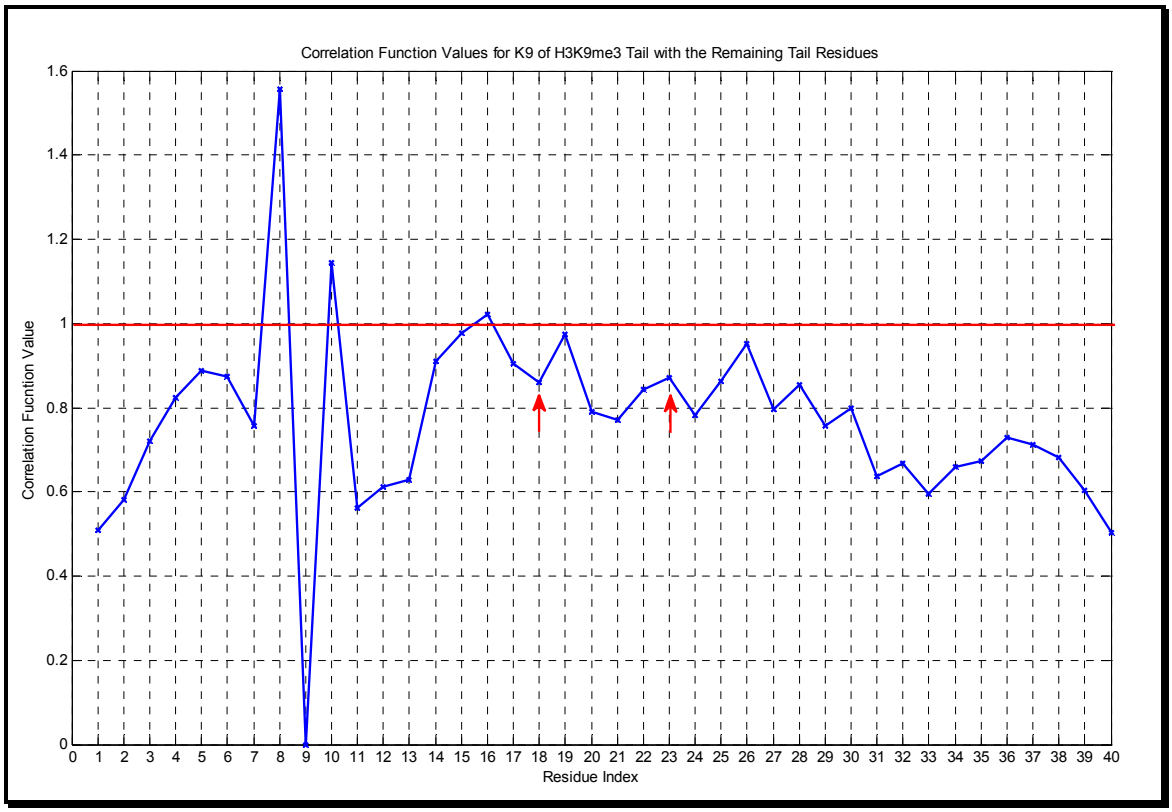


**Figure 6.12.** $g^o(i,9)$ Values for K9 of the H3K9me3 tail

Furthermore, the methylation of K9 was proven to prevent the acetylation through the H3 tail [7]. Table 6.4. and Figure 6.12. obviously displays that the two acetylation sites, K18 and K23, are inhibited by this modification besides K14 that is previously shown to be impeded by the tri-methylation of K9. All together these cooperativity results agree well with the experimental evidences, confirming the efficiency of the techniques followed in the computations.

For the investigation of the interactions between the residues of K4-K9, K4-K14, K4-K18, K4-K23, K9-K14, K9-K18 and K9-K23 pairs, $g^o(i,j)$ computations were carried

out in the case of unmodified H3 tail. Table 6.5. below displays the results for these computations.

**Table 6.5.** Correlation function values computed for the unmodified H3 tail.

| | Unmodified H3 Tail | | |
|---|---|---|---|
| **Pairs** | $P(j)$ | $P(i,j)$ | $g^o(i,j)$ |
| **(K4, K9)** | 0.0260 | 6.3203e-04 | **0.9373** |
| **(K4, K14)** | 0.0260 | 8.4276e-04 | **1.2498** |
| **(K4, K18)** | 0.0260 | 7.3642e-04 | **1.0921** |
| **(K4, K23)** | 0.0260 | 8.1031e-04 | **1.2017** |
| **(K9, K4)** | 0.0250 | 6.3203e-04 | **1.0085** |
| **(K9, K14)** | 0.0250 | 8.2111e-04 | **1.3102** |
| **(K9, K18)** | 0.0250 | 7.2005e-04 | **1.1490** |
| **(K9, K23)** | 0.0250 | 7.7063e-04 | **1.2297** |

According to the unmodified H3 tail results displayed in Table 6.5., the K4-K9 residue pair has negative cooperativity with a $g^o(i,j)$ value smaller than 1, while the remaining residue pairs have positive cooperativity with $g^o(i,j)$ values greater than 1. The correlation function results for K4-K9 and K4-K14 pairs show consistency with the results obtained for the pairs from the H3K4me3 case. There is a moderate difference between the results of K4-K9 pair, the H3K4me3 case result having a smaller value, indicating that the addition of methyl groups increased the degree of negative cooperativity. However, the difference between the unmodified tail and H3K4me3 cases for the K4-K14 pair is insignificant demonstrating that the positive cooperativity information between these residues also exists in the absence of methyl groups. Similar results obtained for K4-K18

and K4-K23 pairs pointing out that the methylation of K4 does not change the acquired cooperativity information between the residues. Unlike these pairs, the K9-K4 and K9-K14 pairs exhibit contrary behavior compared to the H3K9me3 case, having $g^{o}(i,j)$ values both greater than 1, thus being positively correlated. These results suggest that methylation of K9 entirely disrupts the cooperativity information between these residues, highlighting the importance of methylation of the $9^{th}$ lysine residue. Also, similar results indicating the significant effects of K9 methylation were obtained for the K9-K18 and K9-K23 pairs.

# Chapter 7

# CONCLUSION AND DISCUSSION

The N-terminal tail of the histone H3 has a crucial role in epigenetic due to the post-translational modifications that the residues are exposed. Especially, the modifications of K4, K9 and K14 residues have proven to be the switch points in cross-regulation patterns which have diverse downstream outcomes. MD simulations of the unmodified H3 tail, the H3K4me3 tail and the H3K9me3 tail were carried out to investigate the cooperativity between the modifications of the three residues mentioned above. Subsequently, the modal analysis was carried out by using MD trajectories for better understanding the fluctuation dynamics of the tails at residue level. The modal analysis results revealed the modification sites on the three H3 tails at the fastest modes, which denotes the effectiveness of the technique for unveiling the binding site residues. For the unmodified H3 tail, all of the residues that are exposed to a modification were observed as regional vertical contours of $\left\langle \Delta R_{ij}^2 \right\rangle$ map. Furthermore, for the H3K4me3 tail, a significant decrease at the $\left\langle \Delta R_{ij}^2 \right\rangle$ correlations of the K9 residue was monitored, whereas for the H3K9me3 tail, the $\left\langle \Delta R_{ij}^2 \right\rangle$ correlations of the K4 residue were observed to increase although the rise is insignificant compared to the correlations of the remaining residues. These observations point out that the modifications of these two residues affect the binding site properties of each other and coincide with the experimental results which suggest that the tri-methylation of K4 and K9 prevents the modifications of one other. Moreover, for better understanding the cooperativity of these residues at theoretical level, the distribution functions were related to

the $\left\langle \Delta R_{ij}^2 \right\rangle$ correlations and the correlation function $g^o(i,j)$ computation was performed for the residue pairs. According to these results, the negative cooperativity between K4-K9 and K9-K14 pairs were obtained, having a $g^o(i,j)$ value smaller than 1, whereas the positive cooperativity with a $g^o(i,j)$ value greater than 1 was obtained for K4-K14 pair. Based on these indications of the three residue pairs, the same correlation function and interaction energy analysis were carried out with the whole H3 tail residues for the H3K4me3 and H3K9me3 cases, in order to predict the cooperativity between the tri-methylated K4 and K9 with the remaining residues. According to this further analysis, the two acetylation sites that are K18 and K23 were shown to be activated by the tri-methylation of K4, whereas inhibited by the modification of K9 as suggested by the experimental studies.

These statistical computations were also carried for the unmodified H3 tail case to investigate the behavior of the aforementioned residues. Based on the comparisons between the methylated cases and the unmodified case, positive cooperativity information for the K4-K14, K4-K18 and K4-K23 pairs and negative cooperativity information for the K4-K9 pair was observed to occur also in the unmodified case. However, the addition of methyl groups to K9 converts the cooperativity information between the K9-K4, K9-K14, K9-K18 and K9-K23 pairs, highlighting the importance of this modification. These results illustrates that K4 methylation does not change the interaction information that occurs between residues in the absence of methyl groups significantly, on the other hand methylation of K9 adjusts this information entirely.

In this study the cooperative act of the histone H3 tail modifications were aimed to be investigated by MD simulations. The modification sites on the histone H3 tail was revealed by the analysis of the $\left\langle \Delta R_{ij}^2 \right\rangle$ correlations at the fastest modes from the MD trajectories. Furthermore, the $\left\langle \Delta R_{ij}^2 \right\rangle$ correlations were associated with the distribution functions for the

computation of the correlation functions, $g^o(i,j)$. From the cooperativity point of view, the inhibition and activation mechanisms of the modifications between the residues was demonstrated to be explained accurately by the investigation of the $g^o(i,j)$, which establishes an intrinsic communication between the H3 tail residues besides the cell signaling pathways. Moreover, the techniques followed in the computations were confirmed to be efficient for the exploration of the cooperativity at the statistical thermodynamics basis.

## BIBLIOGRAPHY

[1] C. L. Peterson, & M. A. Laniel, Histones and histone modifications, Curr. Biol. 14 (2004), R546-51.

[2] J. I. Loizou, et al., Epigenetic information in chromatin: the code of entry for DNA repair, Cell Cycle 5 (2006), 696-701.

[3] J. A. Latham, & S. Y. Dent, Cross-regulation of histone modifications, Nat. Struct. Mol. Biol. 14 (2007), 1017-1024.

[4] B. D. Strahl, & C. D. Allis, The language of covalent histone modifications, Nature 403 (2000), 41-5.

[5] M. Biel, V. Wascholowski, & A. Giannis, Epigenetics-an epicenter of gene regulation: histones and histone-modifying enzymes, Angew Chem. Int. Ed. Engl 44 (2005), 3186-216.

[6] C. D. Carvin, & M. P. Kladde, Effectors of lysine 4 methylation of histone H3 in Saccharomyces cerevisiae are negative regulators of PHO5 and GAL1-10, J Biol. Chem. 279 (2004), 33057-62.

[7] W. Fischle, Y. Wang, & C. D. Allis, Histone and chromatin cross-talk, Curr. Opin. Cell Biol. 15 (2003), 172-83.

[8] H. B. Callen, Thermodynamics and an introduction to thermostatistics, Wiley, New York, 1985.

[9] O. N. Yogurtcu, M. Gur, & B. Erman, Statistical thermodynamics of residue fluctuations in native proteins, J. Chem. Phys. 130 (2009), 095103.

[10] T. Haliloglu, and B. Erman, Analysis of correlations between energy and residue fluctuations in native proteins and determination of specific sites for binding, Phys. Rev. Lett. 102 (2009), 088103.

[11] T. Haliloglu, E. Seyrek, and B. Erman, Prediction of binding sites in receptor-ligand complexes with the Gaussian Network Model, Phys. Rev. Lett. 100 (2008), 228102.

[12] A. Ben-Naim, Statistical Thermodynamics for Chemists and Biochemists, Plenum Press, New York, 1992.

[13] C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder, and T. J. Richmond, Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution, J. Mol. Biol. 319 (2002), 1097-113.

[14] Accelrys Software Inc., Discovery Studio Visualizer, Release 2.0, San Diego: Accelrys Software Inc., 2007.

[15] Duan, Y. et al. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J. Comput. Chem. 24 (2003), 1999-2012.

[16] http://q4md-forcefieldtools.org/Tutorial/Tutorial-1.php.

[17] Gaussian 03, Revision C.02, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M.

Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, Gaussian, Inc., Wallingford CT, 2004.

[18] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, Scalable molecular dynamics with NAMD, J. Comp. Chem., 26 (2005), 1781-1802.