ABSTRACT

Title of dissertation: USING ENERGY LANDSCAPE THEORY TO UNCOVER THE ORGANIZATION OF CONFORMATIONAL SPACE OF PROTEINS IN THEIR NATIVE STATES.

Davit A Potoyan, Doctor of Philosophy, 2012

Dissertation directed by: Professor Garegin A Papoian
Chemical Physics Program

The functional motions of proteins navigate on rugged energy landscapes. Hence, mapping of these multidimensional landscapes into lower dimensional manifolds is imperative for gaining deeper insights into the functional dynamics. In the present work we implement novel computational schemes and means of analysis to characterize the topography of conformational space of selected proteins and also to elucidate their functional implications. The present thesis is divided into two parts, where we focus on the case studies of the intrinsically disordered histone tails and the representative allosteric protein Adenlyate Kinase. In particular, analyzing the energy landscapes of histone tails, we find preferential clustering of transient secondary structural elements in the conformational ensembles, which have a dramatic impact on the chain statistics, conformational dynamics and the binding pathways. In the study of Adenylate Kinase we use a novel nonlinear order parameter to rigorously estimate the free energy difference between allosteric states and map out the plausible pathway of transition, which reveals important structural and thermodynamic

insights about the mechanism of allostery in Adenylate Kinase. Taken together our findings indicate that the organization of conformational space of functional proteins is delicately crafted to ensure efficient functional regulation and robust response to external signals.

USING ENERGY LANDSCAPE THEORY TO UNCOVER THE ORGANIZATION OF CONFORMATIONAL SPACE OF PROTEINS IN THEIR NATIVE STATES.

by

Davit Potoyan

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2012

Advisory Committee:
Professor Garegin Papoian, Chair/Advisor
Professor Christopher Jarzynski
Professor John D. Weeks
Professor Vitali Tugarinov
Professor Maria Cameroon
Professor Sergei Sukharev, Dean's Representative

To my dearest mother Marietta and my family.

Acknowledgments

I find myself highly fortunate and honored for being surrounded by so many wonderful people in the form of my family, friends and colleagues, who have given me their love and support and to them I completely owe the body of work in this thesis.

I would first like to thank my research adviser, Dr. Garegin Papoian for accepting me in his research group and providing guidance, support and critical analysis on all aspects of my work. I am also grateful to him for giving me the freedom to occasionally wonder on my own and wrestle with the new and exciting ideas, some of which turned out to be successful. For me, Garyk has been a colleague, teacher and a friend at the same time, which initially used to be quite confusing in terms of finding out which one was really talking to me.

His advices about science and life have played an important role in shaping me both as a person and as a scientist. One of the most important lessons that I have learned from Garyk is his problem-centric doctrine. According to it, when an interesting problem is found one needs to dive in it voraciously, developing all the necessary tools, concepts and ideas for its solution, instead of just looking for the vulnerable problems, for which there happens to be an established way of study. This advice had a profound influence on me and taught me not to be content with knowing some narrow subfield in which I happen to specialize, but to traverse the learning barriers with confidence and explore the landscape of science until I find my niche in it.

I would like to express my gratitude to my friends and lab-mates, both past and present who for all these years not only endured me with patience but showed earnest enthusiasm for time and again in helping me out on various occasions. Specifically, I would like to mention, Mr Natsuki Tanaka for showing me the zen of scripting, Ms Maria Minakova for being my "second mom" in the lab, Dr Pavel Zhravlev for teaching me the real russian style of doing stuff, Dr Konstantin Popov for making life in the lab more enjoyable, Drs Sangwook Wu, Longhua Hu and Alex Savelyev for their valuable advice, Mr Chris Materese and Mr David Winogradoff for their comradeship and for checking my manuscripts and helping me to further improve my "purfect" english. I am also thankful for my outstanding teachers who set an example to strive for and who, in the good traditions of science, gave their best to educate their younger colleagues. In that regard I would like to single out professors Michael Rubinstein and Christopher Jarzynski, whose rigorous and thoughtful approach to science has been a source of inspiration for me.

I would also like to say a few words of appreciation to some of my fellow university friends, namely to Alexandra Kuvaeva, Timur Sultanov, Dilyara Yusupova, Denis Filimonov, Jaime Gomez, Nof Abuzainab, Cecilia Rorai, Leysan Khakimova, Zulya Tomova, Natalya Dikhanova, George Zaki, Kleoniki Vlachou, Raef Bassily and Mike Azatov who added much needed joy to my life and who played a major role in the preservation of my sanity after all these years in the graduate school. I am also thankful to my crazy armenian friends from North Carolina who eased my transition into this continent, and with whom I shared many days filled with joy, laughter and unheard of optimism. The staff at the Chemical Physics program

made my life so much more enjoyable. Without Mrs Debbie Jenkins with her caring personality, and Prof Coplan with his warm and fatherly attention towards his students, there would have been a sizable void in my graduate life.

At last but by no means least I would like to thank my mom and the members of my extended family who even in the most difficult of times went at great length to help me achieve my aspirations. I hope one day I will live up to the great expectations you have for me and will earn the love and care that I have received from you.

Table of Contents

List of Figures

List of Abbreviations

ADK	Adenylate Kinase
AMBER	Assisted Model Building with Energy Refinement
CHARMM	Chemistry at HARvard Macromolecular Mechanics
ff99SB	An amber parameter set
H4, H3, H2B, H2A	Names of histone proteins
IDP	Intrinsically Disordered Protein
kcal	Kilocalories
LAMMPS	Large-scale Atomic/Molecular Massively Parallel Simulator
mM	millimolar, 10^{-3} mol/L
MC	Monte Carlo
MD	Molecular Dynamics
NAMD	NAnoscale Molecular Dynamics
NMR	Nuclear Magnetic Resonance
NTP	Isochoric-Isoenergetic ensemble
NVE	Isothermal-Isobaric ensemble
NVT	Isothermal-Isochoric ensemble
PC	Principal Component
PCA	Principal Component Analysis
dPCA	Dihedral Principal Component Analysis
PDB	Protein Data Bank
PMF	Potential of Mean Force
REMD	Replica Exchange Molecular Dynamics
RMSD	Root-mean-square deviation
TIP3P	3 site water model
US	Umbrella Sampling
WHAM	Weighted Histogram Analysis Method
WT	Wild Type

Chapter 1

Introduction and Preview of Thesis

1.1 The Chromatin Problem

The completion of the human genome in 2001 [1] marked the beginning of a post-genomic era, largely shifting the research focus from the linear sequence analysis to understanding the intricacies of structural organization and regulatory mechanisms of the genome [2–4]. The latter undertaking demands a thorough scrutiny of chromatin: a functional form of the genetic material consisting of DNA in a tight association with a group of proteins. As is well known, the sequence information of the DNA is self sufficient for fully defining the chemical makeup of virtually every cellular component. However, for the genetic information to be of any value, it needs to be managed by the whole cell to tailor its varying needs. In other words, on top of the static sequence information there has to be some set of dynamical regulatory instructions, which upon necessity will grant or restrict the access to certain portions of the genome for transcriptional readout, recombination and repair. The management of the genome is handled by a sophisticated molecular machinery of chromatin, that maintains and regulates the informational flow in cells with a high fidelity [5, 131]. Understanding the physical underpinnings of the self assembly and dynamical regulation of the genome constitutes a broad problem of contemporary research interest, which requires a highly interdisciplinary approach.

During the last two decades the chromatin has been under intensive investigation mostly by structural biologists and biochemists, who have made important advances in terms of clarifying the nature of key protein complexes [12], providing clues about the different levels of chromatin organization [57, 68, 129] and discovering novel correlations of gene activity with post-translational chemical perturbations [15, 66]. However, without diminishing the tremendous insight that has been gleaned from these studies, the physical mechanisms of DNA folding and functional regulation to this day remain somewhat enigmatic [14]. The molecular complexity of the system and length scales that are involved make the problems related to chromatin quite challenging, because of which it has been to a significant extent unexplored by the chemical physics community.

The present work is an attempt to provide a molecular level explanation for the behavior of one of the key mediators and regulators of the chromatin, known as the histone tails, using the computational tools and concepts of the equilibrium statistical mechanics. Histone tails, as we will see shortly after, play a prominent role in the DNA folding and regulation of the genetic processes. Understanding the molecular basis of their behavior is an important step towards understanding the organization and functionality of the chromatin. We now proceed to outline in general terms the current state of knowledge about chromatin, highlighting the crucial role played by the histone tails.

As our intuition tells, for the valuable information to be useful, it has to be stored in an organized fashion and in a "safe place" just like books in the library or computer programs on hard disks. It is not surprising, therefore, that the idea of a

hierarchically organized genetic material has been proposed even before the elucidation of the first canonical structure of the DNA. The celebrated example of one such theoretical musing is the Schrodinger's insightful analogy of the DNA with an aperiodic crystal [46], which implies an orderly arrangement created by the non-repeating patterns of genetic units. From the biological point of view there are fundamental reasons [5] why the genome needs to be well organized. Firstly, the genomic DNA has to fit within the boundaries of the nucleus, which provides protection of the all important informational content from the deleterious cytosolic environment. Secondly, the DNA in the nucleus needs to be compacted in a structured way, which would allow timely access of the various portions of the DNA for transcriptional readout, manipulation and repair [5]. These observations hint at an existence of compact, but yet highly organized structure, also known as the chromatin.

Quantitatively speaking, the genome of an eukaryotic cell spans several dozens of billions of base pairs, which are distributed among the chromosomes, in sum equating to $\sim$1m long stretch of the DNA, stored in the micrometer sized nucleus [16]. The physical basis and the biological implications for this nearly million fold compression of highly charged semi-flexible (persistence length $l_p \approx 50nm$) macromolecule are not well understood and are currently under active investigation. Nevertheless, the most basic levels of organization are now well established and currently the novel experimental techniques, such as the Chromatin Conformation Capture [20] and its sequels [22], are beginning to show glimpses of the higher levels of the hierarchy [21]. At the most basic level, the compaction of genome starts with a $\sim$146bp stretch of double stranded canonical DNA which is wrapped around the spool of the histone

octamer by ∼1.6 superhelical turns. This complex of histone octamer with the DNA is called nucleosome (see Fig. 1.2), which is the fundamental repeating unit of the chromatin. The histone octamer consists of four types of proteins named as histones H4, H3, H2B and H2A . The histones are highly basic multi-domain proteins, which contain two structurally distinct units: a structured globular or core part and the disordered N- or C- terminal segments which are commonly called the histone tails (Fig. 1.2). The globular part of the histone makes numerous contacts with the DNA periphery, which compensates with an excess the large energetic price for the sharp bending of the DNA strands [131]. Rough estimate using the worm like chain model for the DNA gives $2 - 3\,k_BT$ of net free energy gain for binding to histone octamer, which is large enough to keep the nucleosomes stable, but small enough for remodeling complexes and binding agents to perturb or entirely decompose the nucleosomal structure. The histone tails, through their numerous contacts with the adjacent nucleosomes further stabilize the DNA-histone association by facilitating the formation of a poly-nucleosomal aggregates.

In the absence of stabilizing forces, the poly-nucleosomal arrays are nothing but a linear string of nucleosomes joined by the short stretches of linker DNA, referred to as a "beads on a string" structure (Fig. 1.1) of the chromatin [80, 191]. These pseudo-linear poly-nucleosomal arrays are only observed during the stretching experiments or by creating solutions with low ionic strength. Under the native cellular conditions the poly-nucleosomal arrays fold into densely packed filaments with a diameter of ∼30nm. The higher levels of organizations are less well characterized, but presumably involve some kind of regular packing of the 30nm fibers

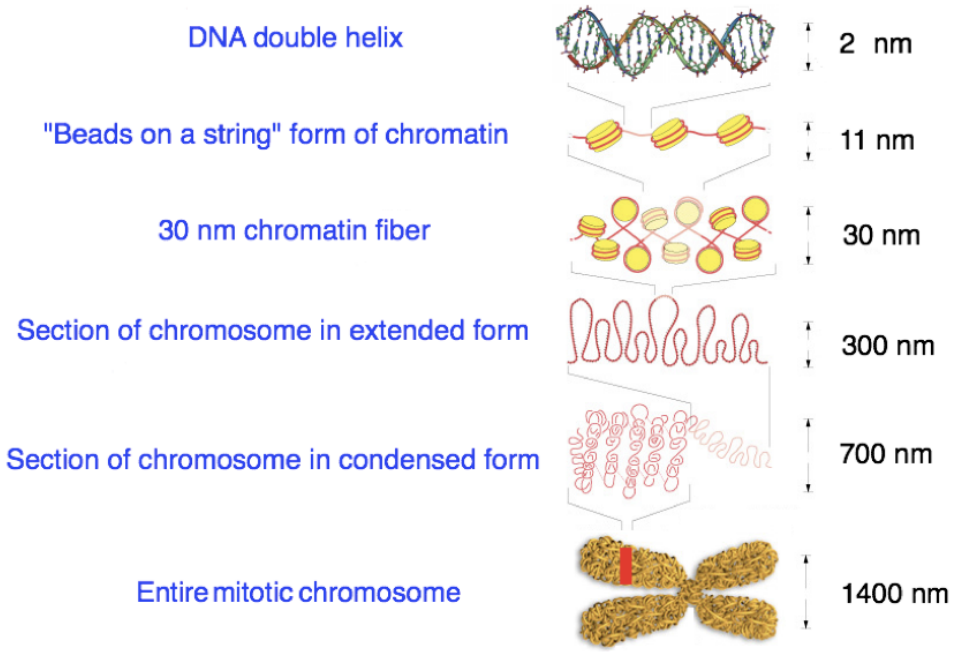

Figure 1.1: Hierarchical organization of DNA as a part of the Chormatin. Copyright 2007 from Molecular Biology of the Cell by Alberts et al. Reproduced by permission of Garland Science/Taylor & Francis Books, Inc.

which form the dense nucleoprotein mesh of the chromosomes (see Fig. 1.1). Obviously this descriptive picture of the chromatin compaction is highly simplified and should be taken with a grain of salt, since at the present there is no consensus on the higher levels of organization of chromatin and even the existence of the 30nm fibers is still hotly debated [10, 14, 191]. Nevertheless, the role of histone tails in the chromatin organization has been clearly established by the experiments with the tailless nucleosomes [9, 80, 170], which showed that in the absence of histone tails poly-nucleosomal fibers lack the propensity to fold into dense chromatin fibers.

Besides their roles as mediators of chromatin's high level structural organization, the histone tails also serve as a platform for gene regulation via a variety of

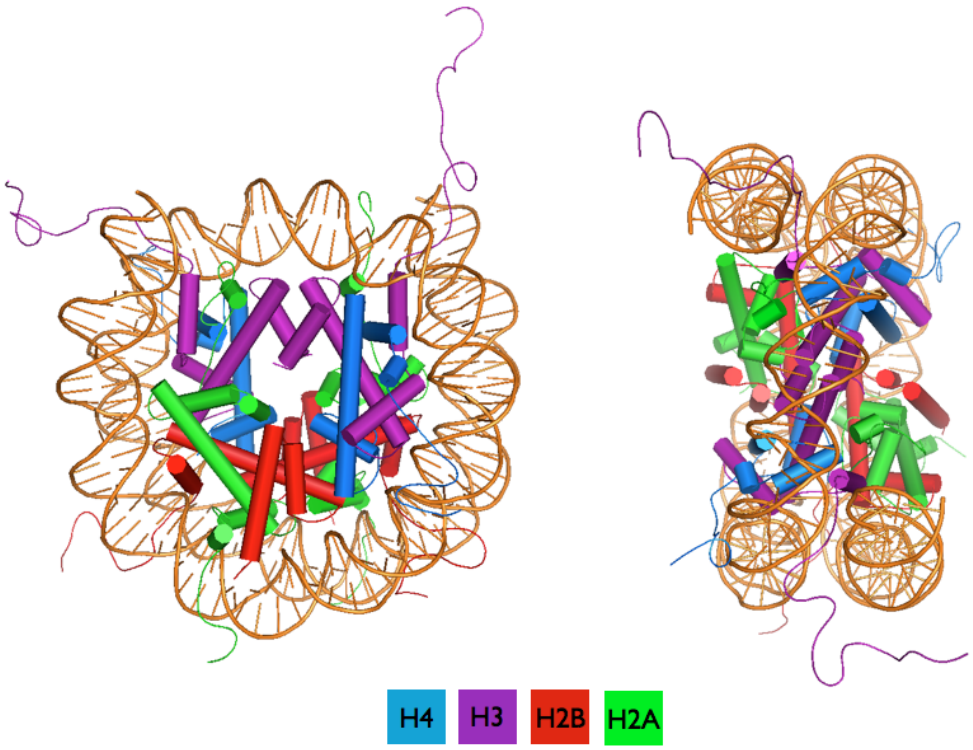

Figure 1.2: Orthogonal views of the crystal structure of the nucleosome at 1.9$\mathring{A}$ resolution (PDB ID: 1KX5). Histones are colored as indicated at the bottom of the figure. The histone tails are seen as flanking disordered segments exiting the nucleosomal core either from the top/bottom (H4 and H2A) or through the helical gyres(H3 and H2B).

post-translational modifications [15,139,176]. These modifications come in different types (such as acetylation, methylation and ubiquitination of charged residues) each of which and also their different combinations carry specific messages for a cell that signals a need to activate or repress diverse genetic activities. These needs are met by two different mechanism: either by recruiting specialized proteins, which bind to histone tails and upon recognition fulfill the their tasks or by directly altering the physical state of chromatin, which either decompresses or further buries certain por-

tions of the of the DNA in the chromatin fibers. The topic of the post-translational modifications is very rich and is full of speculations and unsolved puzzles [15,107,168] which is colorfully summarized by the notion of a "histone code" [139,173] claiming that every type of modification contains a unique code that is being recognized and responded by the cellular machinery. The idea of a 'histone code' turned out to be somewhat controversial [12, 15], however the fact that many of modifications, either by direct or indirect means, induce changes of the chromatin's functional and structural state is now proven beyond any reasonable doubt [15, 168]

In this work we are interested in understanding the modifications that proceed via direct physical mechanisms, of which the acetylation of the H4 tail is the most pronounced case and is regarded as the paradigm for transcriptional activation [13,177,186]. The currently accepted qualitative picture behind the acetylation induced transcriptional activation is the following [8, 13, 135]: when an acetyl group is attached on the H4 tail, it causes drastic some changes in the tail-nucleosome interactions, which then leads to local unwrapping of the chromatin fiber, exposing the underlying DNA for transcriptional agents [5]. One indirect evidence for this picture comes from the fact that levels of gene expression are well correlated with the degree of acetylation. Additionally in the landmark experiment of Shogren-knaak et al [135] it was demonstrated that mono-acetylation of LYS-16 on the H4 tail completely unravels the 30nm chromatin fibers, thus providing the first direct structural evidence for the disruptive effect of H4 tail acetylation. This qualitative picture is nonetheless largely incomplete, because it defines the causes and effects of acetylation, while lacking the molecular level explanation for the observed behavior.

In particular the mechanism by which the acetylated H4 disrupts the chromatin fiber is not known.

In another set of experiments [80] the disruptive effect of histone tail acetylation was shown to be reversible by increasing the ionic strength of the solution, which worked for all histone tails with the exception of the H4 tail. These experiments, show the unique role of H4 tail in stabilizing chromatin fiber and also hint that the driving force behind the observed behavior might be non-electrostatic in nature [80]. At last we note that all of the mentioned issues are related and form one of the central questions of chromatin science which has been puzzling scientist for over a decade. In the chapter III of this thesis we propose a molecular level mechanism that sheds light on this issue.

In the crystal structure of the mono nucleosome histone tails are seen as disordered coils [7] which has prompted the proposition that tails function by a simple electrostatic mechanism by binding non-specifically to the DNA and alleviating the repulsive intra-nucleosomal interactions [80, 160]. We note however that this view of the histone tails is incapable of explaining the high level of sequence homology in histones [49], which strongly implies a unique functionality. Unfortunately at present the issues regarding the degree of order in the conformations of histone tails, their roles in chromatin organization and the molecular explanations for functional behavior are still unresolved. The study of structural and dynamical aspects of histone tails has proven to be quite challenging for multiple reasons. Firstly from the experiments it is challenging to get structural snapshots of a highly dynamical entities such as histone tails which furthermore under *in vivo* conditions operate at

a constantly changing molecular environment. Secondly, and on a more fundamental level, our general knowledge regarding the nature of the intrinsically disordered proteins and how the disorder helps in achieving the observed balance of functional diversity and specificity is incomplete.

In the subsequent chapters of the thesis I mainly focus on the understanding the physicochemical properties of histone tails and their functioning mechanisms. Using the well established computational tools of the equilibrium statistical physics (REMD, US, the free energy techniques) I address the specific questions concerning the structural order in the conformational ensemble of histone tails, their role in the DNA binding and their functional implications.

1.2 Computing Free Energies of Allosteric Transitions in Proteins

Computing the free energies of condensed matter systems is a laborious task, usually requiring long runs of MD or MC simulations [91, 92]. The main source of difficulty lies in the complicated form of the highly multidimensional potential energy function $U(r^N)$, which creates a rugged free energy landscapes. Such landscapes pose a significant computational burden for the straightforward sampling techniques, such as the MD or Metropolis' MC. Although, knowing the underlying physics of the problem and hence having a rough idea about the features of the free energy landscape, one may devise methods that take advantage of the specificities of the system in order to simplify the calculations. This strategy seems to have been successful, as there is now a large arsenal of techniques [24, 39, 91, 92] for tackling the specific class of problems, such as solid state transformations [25], phase transitions [28, 92], ligand macromolecule binding [26, 27], conformational transitions in proteins [29–31, 220, 238] etc. A common procedure for computing free energy differences in computer simulation consists of choosing a reaction coordinate or some progress variable connecting the reference states and performing a simulation which "transforms" one state into another. To serve its purpose, the reaction coordinate must first of all clearly discriminate the states A and B on the structural grounds. Ideally, one may also infer relevant kinetic information if the states are connected in a way which captures the energetic bottlenecks (saddle points) of the transition. However, the latter goal is more challenging and is beyond the scope of the present work.

The transformation between states A and B is typically done by systematically varying the path coordinate, which is often accompanied by restraining it at an intermediate values to acquire better statistics. For the latter task there are number of popular techniques, such as the free energy perturbation [38, 92], umbrella sampling [34] and blue moon sampling [37, 39]. Thus, if we for the moment assume the sufficient level of sampling, the only inevitable intellectual challenge for the free energy calculation would be in finding a good reaction coordinate. Unfortunately, for many biophysical problems finding such a coordinate can be a fairly daunting task. One such case where finding of a good reaction coordinate is problematic is the functional transitions of proteins or the allosteric transitions, which are conformational changes triggered by the binding of a smaller ligand.

To illustrate the difficulties in studying the allosteric transition it is instructive to sketch the qualitative picture of the protein folding free energy landscape (see Fig. 1.3a) and contrast the allosteric transition with the protein folding. The folding transition on the landscape can be schematically represented as a downward movement of a putative order parameter along the funnel [17, 75, 108]. The low values of the order parameter correspond to disordered coils and the high values are associated with the native functional states of the protein. As is well known, the folding process is well correlated with formation of native contacts which induces a distinct conformational change from an unstructured random coil into a compact globule. Therefore, a reaction coordinate that quantifies the number of native contacts will be well suited for computing the free energy difference as it clearly discriminates the reference states(folded and unfolded) and also correlates with the transformation.

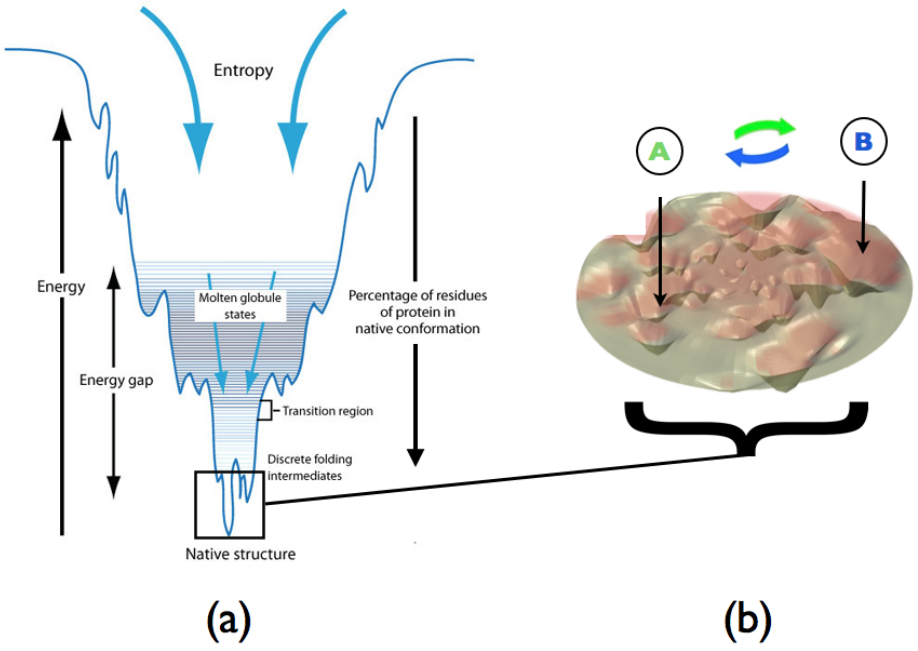

Figure 1.3: (a) The schematic folding free energy funnel. (b) The higher resolution picture of the native state showing a presumed conformational transition path between two allosteric states. Figure (a) is reproduced with modifications from the educational site *www.Learner.org*. Figure (b) is reproduced with the permission of author (G. A. Papoian Proc. Natl. Acad. Sci. USA 2008, 105, 14237)

The role of such coordinate can play the Q value [17, 18], the RMSD (Root-Mean Square Deviation) with respect to the native structure [23] and also sometimes the radius of gyration [19].

The allosteric transition on the same diagram would be a thermally activated barrier crossing at the very bottom of the free energy landscape (see Fig. 1.3b). This kind of transition involves subtle structural rearrangements, for which there is no immediately obvious geometric parameter that captures the key features of the transformation. Thus, the coordinates which are well suited for the folding can no

longer be applied to allosteric transition (see [238] for more details) since they do not provide sufficient level of structural resolution and they do not correlate with the transition. In the chapter III of this thesis we present a method for estimating the free energies of allosteric transitions using a nonlinear reaction coordinate that overcomes the listed difficulties. We apply our technique on the transformation between open and closed states of the E coli Adenylate kinase (ADK). We show, that performing umbrella sampling simulation on our coordinate leads to a robust free energy difference between allosteric states of ADK.

1.2.1 Umbrella Sampling

The central aim of the umbrella sampling (US) is to accelerate the sampling on a transformation path by replicating the system among multiple windows and restraining (by adding a biasing potential to a reference hamiltonian) them at different intermediate points along the path [34, 36]. The strength of the method lies in the fact that the added potentials bias the system to sample regions, which would otherwise be rarely visited during the conventional MD simulation for the same amount of cumulative time. However, the gain in sampling comes with a price since the introduction of the biasing potential alters the natural course of the dynamics. Fortunately the clever unbiasing trick can fully recover the underlying thermodynamics, given the assumption of ergodicity and sufficient sampling in all of the windows. The functional form of restraining potential is arbitrary, but the harmonic form $w(\xi(r)) = k(\xi(r) - \xi)^2$ has become widespread in the applications because of its simplicity and intuitive appeal. By the judicious choice of the spring constants k_i and strategically positioned windows, the umbrella sampling thus aspires to provide a quicker way of inferring free energy by running relatively short set of Molecular dynamics or MC simulations in the Independent windows. In more practical terms, by performing umbrella sampling, we essentially attempt to extract the unbiased distribution function (given by Eq. 1.1) of our reaction coordinate ξ, which is a relatively slowly converging function in the conventional MD simulations.

$$P^0(\xi) = \frac{1}{Z_0} \int \mathrm{d}r^N \, e^{-\beta U(r)} \, \delta(\xi(r) - \xi) \tag{1.1}$$

Adding the biased potentials (at $i = 1, ..N$ points along the path) to the reference potential energy function $U_i^b(r) = U_0(r) + w_i(\xi(r))$, one samples the corresponding biased probability distribution given by:

$$P_i^b(\xi) = \frac{1}{Z_i} \int \mathrm{d}r^N \, e^{-\beta(U(r)+w_i(\xi(r))} \, \delta(\xi(r) - \xi) \tag{1.2}$$

The delta function in the argument makes the bias Independent of r^N, and it can be factored out of the integral resulting in

$$P_i^b(\xi) = \frac{e^{-\beta w_i(\xi)}}{Z_i} \int \mathrm{d}r^N \, e^{-\beta U(r)} \, \delta(\xi(r) - \xi) \tag{1.3}$$

Recognizing the equation for the unbiased probability distribution in the and doing the rearrangement of terms we obtain,

$$P^0(\xi) = \frac{Z_i}{Z_0} e^{\beta w_i(\xi(r))} P_i^b(\xi) = e^{-\beta(F_i - F_0)} \, e^{i\beta w_i(\xi(r))} P_i^b(\xi) \tag{1.4}$$

The last expression provides a direct route to recover the free energy of the original unbiased system from the biased probability distribution. For every window i we get the appropriate parts of the full potential of mean force on our reaction path by the virtue of the following expression:

$$A_i(\xi) = -\frac{1}{\beta} ln \, P_i^b(\xi) - w_i(\xi) + F_i \tag{1.5}$$

The Eq. 1.5 is formally exact, although it contains undetermined coefficients in the form of $F_i = F_0 - \frac{1}{\beta} ln \, \frac{Z_i}{Z_0}$ which makes its usage somewhat impractical. In the next section we will see that the determination of F_i coefficients requires the application of special iterative techniques. Once these coefficients are known the free energies in each window can be combined yielding one continuous potential of mean force along

the coordinate. The one obvious drawback of the umbrella sampling is the need to guess k_i coefficients ($w(\xi(r)) = k_i(\xi(r) - \xi)^2$), which depending on the problem can sometimes be a non-trivial task. In general the reference points of biasing potential are placed non-uniformly with more windows near the high barriers and less windows in shallow regions of the underlying landscape. The advantage of the US over the other enhanced sampling techniques is that the simulations are run Independently and in parallel, hence the additional windows can be inserted on the fly by simply monitoring the histogram overlap of the existing ones. At last, I would like to reiterate the common wisdom that the ultimate usefulness of obtained free energies depends on the meaningful choice of the reaction coordinate, which is a separate problem to which we turn in the chapter IV of this thesis.

1.2.2 Techniques for Analyzing Biased Simulations

It has been recognized a long while ago, that a simulation which has been conducted at one condition, can in principle be used for predicting the properties of the system at different condition. To see explicitly how this can be done, consider a constant temperature simulation of a certain physical system characterized by an energy function $U(r)$. In the limit of an infinite MD run, the configurations of the system will be observed with a frequency given by the canonical distribution:

$$P_\beta(U) = \frac{\Omega(U)\, e^{-\beta U}}{Z_\beta} \tag{1.6}$$

Where the $\Omega(U)$ is the classical density of states (implying that the $\Omega(U)\Delta U$ is the number of realizable micro-states in the $[-\Delta U/2, \Delta U/2]$ energy range of a macro-state). Using the distribution at temperature β one can as well infer the distribution at the temperature β', which can likewise be written as:

$$P_{\beta'}(U) = \frac{\Omega(U)\, e^{-\beta' U}}{Z'_\beta} \tag{1.7}$$

Since the density of states $\Omega(U)$ does not depend on the temperature, the last two equations can be combined, yielding the following expression for the new distribution

$$P_{\beta'}(U) = \frac{Z_\beta}{Z_{\beta'}} P_\beta(U)\, e^{-U(\beta' - \beta)} \tag{1.8}$$

This quite general technique, which attempts to map a one set of conditions onto another is known as a single histogram reweighting method [35]. The name for the technique comes from the procedure of reweighting the histograms of the reference simulation (P_β in the Eq. 1.8) for obtaining the probability distributions at the

other states ($P_{\beta'}$ in the Eq. 1.8). Although exact in the limit of infinite sampling, the single histogram re-weighting has very little usage in practice. This is mainly because the $P_\beta(U)$ functions are sharply peaked(effectively delta functions for very large systems), which for most of the values of β do not overlap and therefore provide sparse statistics at the tails of their distributions, making the guesses based on Eq 1.8 highly unreliable.

A good way to increase the reliability of predicting the target state is to conduct multiple simulations, attempting to cover all the relevant regions of phase space. This can be straightforwardly done by an umbrella sampling technique discussed in the last section. In the case of multiple simulations one now faces the problem of utilizing all the data in an efficient way, which will give the most reliable estimate for the target state. Ferrenberg and Swendsen [89–91] proposed one such technique, which combines the histograms from multiple simulations in an optimal way by minimizing the inherent statistical error in each histogram. The method is known as WHAM, which stands for the Weighted Histogram Analysis Method. We provide a brief outline of the derivation [89, 92] of WHAM equations in the context of analyzing the umbrella sampling simulations, which will help us understand its general usage an limitations in the later sections. Suppose we carry out a set of simulations (windows) with the biasing functions $w_i(\xi)$, where the index i labels the simulation (i=1,2...r). The goal is to provide the best estimate of the unbiased probability distribution given by the Eq. 1.1. Each of the biased simulations can be used for estimating the distribution $P(\xi)$ via the Eq. 1.8. However as we discussed in the beginning of this section such estimates are not credible, therefore we use all

of the windows at the same time by writing an estimate as a linear combination of estimates from each window. The weights are given by their overlap under the given condition,

$$P^0(\xi) = \sum_{i=1}^{r} C_i(\xi) P_i^0 = \sum_{i=1}^{r} C_i(\xi) \frac{Z_i}{Z_0} e^{-\beta w_i(\xi)} P_i^b(\xi) \tag{1.9}$$

where the P_i^0 denotes the estimate of P^0 provided by the ith window (or simulation). The central idea behind WHAM is to minimize the standard error $\sigma^2[P^0(\xi)] = \overline{[P_0(\xi) - \overline{P_0(\xi)]^2}}$. The bar over the symbols indicates the mean or the expected distribution. Since the distributions in separate windows are Independent the statistical error turns into a simple sum of errors per window.

$$\sigma^2[P^0(\xi)] = \sum_{i=1}^{r} C_i^2(\xi) \left(\frac{Z_i}{Z_0}\right)^2 e^{-2\beta w_i(\xi)} \sigma^2[P_i^b(\xi)] \tag{1.10}$$

The individual errors are due to the finite sample histogramming, because of which we have $\sigma^2[P_i^b(\xi)] = \sigma^2[N_i(\xi)]/n_i^2 = g_i \langle N_i(\xi)\rangle / n_i^2$. The last expression follows from the standard treatments of the error for the correlated dynamical variables [32, 33]. The $N_i(\xi)$ is the number of counts in the bin $\xi, \xi + \Delta$ and n_i is the total number of counts (simulation length) in the ith window. The $\langle N_i(\xi)\rangle$ is the ensemble average of histogram counts over all the simulations $i = 1, ..r$. The $g_i = 1 + \tau_i$ quantifies the deviation of numerical average from the true distribution and the parameter τ is the integrated correlation function for the ith simulation (see [33]). In the MD simulations one typically assumes that g_i's are all equal to ~ 1, which is tantamount to assuming an equal "quality" of sampling in each window. For the latter to be true, the time interval between recorded values has to be greater than the correlation time in all windows, which will then result in the Independent and uncorrelated

samples. After setting the $g_i = 1$ we are left with a poisson distribution of ξ values in each window $\sigma^2[N_i(\xi)]/n_i = \langle N_i(\xi)\rangle/n_i^2 = P_i(\xi)/n_i$, which we substitute back into Eq. 1.10

$$\sigma^2[P^0(\xi)] = \sum_{i=1}^{r} C_i^2(\xi) \left(\frac{Z_i}{Z_0}\right)^2 e^{-2\beta w_i(\xi)} \frac{P_i(\xi)}{n_i} = P_0(\xi) \sum_{i=1}^{r} C_i^2(\xi) \frac{Z_i}{Z_0} e^{-\beta w_i(\xi)} \frac{1}{n_i} \quad (1.11)$$

In the second part of the equation we used biasing function on a true distribution to get the biased distribution. This step is an approximation of the WHAM, which assumes that the biased distribution can be obtained by simply applying the biasing factor to the true distribution:

$$P_i(\xi) = P_0(\xi) \frac{Z_0}{Z_i} e^{\beta w_i(\xi)} \quad (1.12)$$

Now we are in a position to minimize the statistical error (Eq. 1.11) with respect to coefficients $C_i(\xi)$, subjected to the constraint $\sum C_i = 1$. Introducing the constraint via a Lagrange multiplier $\lambda(\sum C_i - 1)$ and taking the derivative with respect to the coefficient we arrive at the following expression

$$\frac{d\sigma^2[P^0]}{dC_i} = 2\, P_0(\xi) C_i(\xi) \frac{Z_i}{Z_0} e^{-w_i(\xi)} \frac{1}{n_i} - \lambda = 0 \quad (1.13)$$

Simple rearrangement yields the following expression for the coefficients

$$C_i(\xi) = \frac{\lambda n_i e^{w_i(\xi)}}{2P_0} \frac{Z_0}{Z_i} \quad (1.14)$$

Using the normalization condition for coefficients we find the undetermined coefficient λ and substituting it back into the Eq 1.14, we arrive at the final expression for the coefficients

$$C_i(\xi) = \frac{n_i e^{\beta w_i(\xi)} \frac{1}{Z_i}}{\sum_{j=1}^{r} n_j e^{\beta w_j(\xi)} \frac{1}{Z_j}} \quad (1.15)$$

The last expression provides the statistical weights for each window:

$$P_0(\xi) = \frac{\sum_{i=1}^{r} N_i(\xi)}{\sum_{j=1}^{r} n_j e^{\beta w_j(\xi)} \frac{Z_0}{Z_j}} = \frac{\sum_{i=1}^{r} N_i(\xi)}{\sum_{j=1}^{r} n_j e^{\beta w_j(\xi)} e^{-\beta(F_0 - F_j)}} \tag{1.16}$$

The last set of equations can not be solved as it is, since it contains the free energies F_i which are not know from the onset. The free energies for F_i can be related to the target distribution function in the following manner:

$$e^{-\beta F_i} = e^{-\beta F_0} \int \mathrm{d}\xi P_0(\xi)\, e^{-\beta w_i(\xi)} \tag{1.17}$$

Discretizing the above equation and plugging into it the expression 1.16 for the P_0, we obtain the central equation of the WHAM:

$$e^{-\beta F_i} = \sum_k \frac{\sum_{i=1}^{r} N_i(\xi_k) e^{-\beta w_i(\xi_k)} \Delta\xi_k}{\sum_{j=1}^{r} n_j e^{\beta w_j(\xi_k)} e^{\beta F_j}} \tag{1.18}$$

The numerical solution of the Eq. 1.18 is obtained by setting all the $F_i = 1$ and iterating the subsequent solutions until the convergence is reached. At last it is worth mentioning that the WHAM technique is quite general and can be used not only for combining the data from the umbrella sampling simulations but also for simulations conducted under different conditions(e.g, the variants of parallel tempering). For the sake of completeness it should also be noted that besides WHAM there are alternative techniques, such as the umbrella integration [40] and the MBAR [41] (Multistate Bennett Acceptance Ratio), which combine simulations via more elaborate means, most importantly by avoiding the histogram binning step. However, the numerical accuracy of these techniques has been found to be only marginally different from the the WHAM.

1.3 Preview of the thesis.

- The chapter II of the thesis is devoted to the understanding of the polymeric aspects of the histone tails as representatives of a broader class of the intrinsically disordered proteins. In the introduction I review the current state of our knowledge about the intrinsically disordered proteins and motivate the study of histone tails. The results section consists of two subsections, the first of which deals with the intrinsic conformational propensities of the histone tails, while the second concerns with the thermal denaturation and polyelectrolyte behavior. The key contribution of the chapter consists in showing how the presence of the transient secondary structural elements in the conformational ensembles of histone tails can have a dramatic impact on the polymer statistics. In particular we show that the three of the histone tails occupy an intermediate niche between random coils and structured globular proteins. Additionally, we show that the signature of the thermal denaturation for the histone tails involves a novel re-entrant transition not encountered for the globular proteins.

- The chapter III deals with the functional aspects of histone tails, taking as a test case the most extensively studied case of the H4 histone tail acetylation. The central aim of the chapter is to dissect the impact of covalent modifications of amino acid residues on the conformational and binding propensities of histone tails. We have found that the nature of the natively disordered ensemble of the H4 tail significantly affects the way the chain folds upon bind-

ing to DNA. In particular, we show how that post-translational modifications of amino acid residues, such as lysine acetylation can alter the degree of collapse and conformational preferences for a free protein and also profoundly impact the binding affinity and pathways for the protein DNA association. Our findings lead us to propose a hypothesis that can potentially account for the celebrated chromatin "fiber loosening effects", which forms the basis of transcriptional activation in the eukaryotes.

- The chapter IV is devoted to the problem of computational estimation of free energies of allosteric transitions in proteins. In the introductory part we briefly review the free energy techniques proposed prior to our work, highlighting the their shortcomings in the application to the allosteric transitions and the way our method overcomes them. The core of the chapter deals with the application of our technique to the allosteric transition of Adenylate Kinase. We find a marginal free energy gap between the two allosteric states which is in a semi-quantitative agreement with the experiments. Our structural analysis of the transition pathway shows late transition state in the interconversion of the closed form into the open form of the ADK. We have found a significantly higher dynamical structural disorder in the domains of the closed state, which is an evidence for the local frustration between intra- and inter- domain contacts. The general mechanism that emerged from our study is formulated as an entropic transfer mechanism of allosteric transition, whereby the transfer of conformational entropy between different thermal modes of the molecule

helps to minimizes the free energy gap of the allosteric sates.

Chapter 2

Energy Landscape Analyses of Disordered Histone Tails Reveal Special Organization of Their Conformational Dynamics

The chapter is based on the published work of the author:

D.A. Potoyan, G. A. Papoian; *J. Am. Chem. Soc.* **133**, 7405-7415 (2011)

2.1 Introduction

All eukaryotic cells face the dilemma of tightly packaging their genomes inside a small nucleus, while also providing timely access to individual genes for transcription and replication, upon receiving internal and external signals. This difficult physical problem is elegantly solved by formation of a chromatin, a nucleoprotein complex consisting of DNA and histone proteins, densely packed together into regular repeating arrays. The fundamental structural unit of the chromatin is the nucleosome [5, 99, 103, 131, 155], a $\sim 146bp$ long DNA segment tightly wrapped around histone octamer, comprised of 4 pairs of histone proteins: H4, H3, H2B and H2A. Although the structure of nucleosome has been resolved at a near atomic resolution [97, 98, 128], and the nucleosome core particle was investigated using molecular simulations [103, 133], the structural information about higher order polynucleosomal arrays is lacking. Nevertheless, it is clear that the histone

proteins play a prominent role in determining chromatin structure and dynamics, where the latter, in turn, influence many cellular processes such as gene expression, silencing and replication. In particular, histone terminal tails mediate internucleosomal attraction and control chromatin conformation through site specific covalent modifications. The latter mechanism is the basis of the so called histone code hypothesis [5, 71, 101, 107, 118, 139], according to which a specific combination of post-translational covalent modifications creates different biochemical responses by switching on or off various gene transcription and other signaling events. Despite their biological significance, molecular details of how histone tails carry out many of these tasks still remain insufficiently clarified, largely because of their intrinsic disorder. To elucidate molecular mechanisms of histone tail functioning, it is necessary to gain deeper understanding of their internal dynamics and conformational preferences. In the present study, we shed light on the nature of the natively disordered ensembles of various histone tails, focusing in particular on the role of transiently populated secondary structure elements. We also explore the role of mobile counterions in modulating histone tail conformational dynamics.

Beside their pivotal roles in determining chromatin structure and dynamics, histone tails are also known for belonging to a special class of proteins that lack stable and densely packed 3D structure *in vivo*. This class of proteins that explores their unstructured nature to achieve functional promiscuity, is known as intrinsically disordered proteins(IDP) [69, 70, 110, 111, 115, 127, 146, 147]. These proteins are distinguished by several features including low hydrophobicity, high net charge and low sequence complexity, which have a combined effect of impeding the formation of

densely packed globular structures. Recently, several experimental [61,106,138] and theoretical [76,144,145,148,149] studies on various ID proteins demonstrated that despite the lack of major hydrophobic interactions, these proteins often do not show random coil statistics as one might have anticipated. Instead most populated states may be rather compact, with a partial presence of local secondary order, fluctuating in absence of strong stabilizing forces. The intrinsically disordered nature of histone

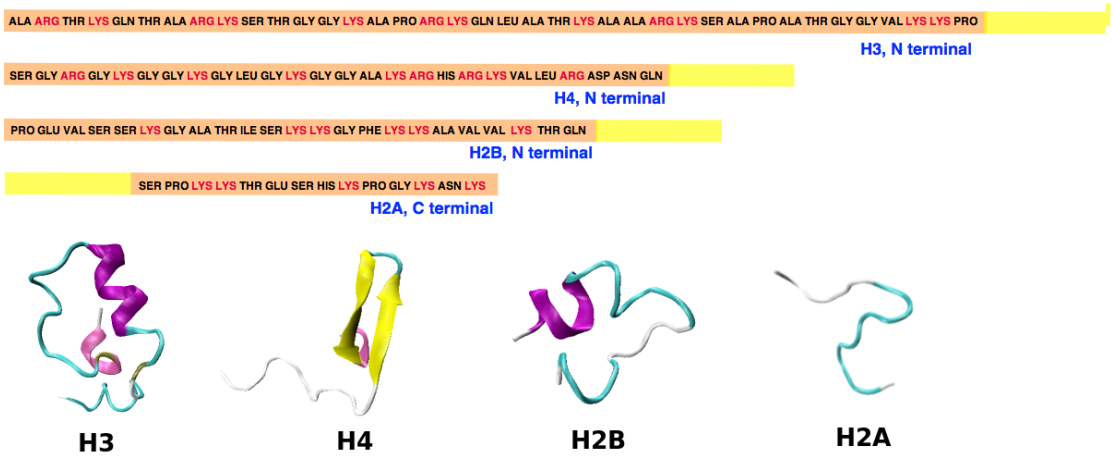

Figure 2.1: Representative conformations of histone tails are shown, obtained from simulations presented in this chapter. Sequences of histone tails are shown in the upper panel, where charged residues are indicated with red and neutral ones with black letters. The solid yellow bars indicate the remaining portions of the histone proteins that are not part of histone tails.

tails became apparent from the X-ray structures of nucleosome, where tail domains appear to sample multiple conformations [97,98]. This high conformational flexibility stems from the amino acid sequences which contain a high number of hydrophobic, charged (1/3 of all residues) and structure breaking (GLY) residues (Fig.2.1). For example $\sim 30\%$ of residues in H3 and H2B tails are hydrophobic, whereas in

H4 and H2A tails only $\sim 15\%$ of residues are hydrophobic. Small globular proteins, on the other hand, are found much higher on the hydrophobicity scale, having sequences which are at least $\sim 50\%$ hydrophobic, with some $\sim 30\%$ of residues consisting of bulky hydrophobic groups (PHE, ILE, LEU) [83, 123]. Histone tail conformational disorder might accelerate binding on-rates through the fly-casting mechanism [111, 134, 143], while their high net positive charges should enhance the binding affinity towards negatively charged DNA surface. The accumulating wealth of experimental data showing specific binding propensities of histone tails eventually led to a suggestion that some tails may adopt specific secondary structures while bound to a linker DNA or acidic patches of core histones [80, 86, 160].

Experiments on nucleosomes using circular dichroism (CD) and combination of hydrogen exchange with NMR showed that H4/H3 tails acquire structured conformations as a part of nucleosome core particle, while H2A and H2B were found to be essentially in random coil like states [55, 86, 153]. The results of CD experiments suggested that some alpha helical structure is present in isolated H4/H3 tails, however, due to the impossibility of selective cleaving of the H3 and H4 tails, distribution of alpha helical amino acids among them had not been assigned. Hence, the possible interpretations allow either helical conformation for most of the H3 tail residues or equal distribution among residues of H4 and H3 tails.

There are only a few atomistic computational studies of histone tails. Of particular relevance to our work is the study of the wild and covalently modified forms of H3 histone tail by Yang, et al. [95]. In the mentioned work implicit solvent Replica Exchange Molecular Dynamics (REMD) simulations were performed finding

that wild type H3 tail populates alpha helical conformations in qualitative agreement with our simulations. In another recent study, Arya et al. [158] found short alpha helical element in the conformations of the H4 histone tail, which is inconsistent with our results. We attribute the discrepancy to the fact that the authors used helically biased forcefields along with the Generalized Born implicit solvent model, which has a combined effect of strongly favoring alpha helices(see [56, 119] and Methods Section). Many coarse-grained studies investigated the role of histone tails in chromatin folding by using low resolution models of histone tails attached to uniformly charged spheres [50–53, 88, 133]. However, for these models to be realistic, should one treat histone tails as random coils, which is a common practice, or as chains with flickering secondary structure elements and potentially complex internal dynamics? The answer to this question should pave a way for large scale computational modeling of chromatin dynamics. In this chapter we are providing the first comprehensive overview of histone tail conformational preferences at high structural resolution.

We carried out all atom REMD simulation of all four histone tails with the aim to clear up the ambiguity related to their structural behavior, to find out the driving force behind observed conformational preferences, and ultimately to devise a suitable qualitative framework for understanding histone tails that can also be utilized in lower resolution studies. The main finding of our study is that three histone tails, H4, H3,and H2B, adopt persistent secondary structural elements and show behavior strongly deviating from random coil statistics. In particular, our results suggest that the H4 tail forms a beta strand in the well known binding region,

while H3 and H2B form alpha helices, consistent with prior experimental findings. In contrast, H2A may be characterized as a random coil. Additionally, we used ideas from the energy landscapes theory [75,108,163,164] and polymer physics [64,122,145] to analyze the behavior of histone tails on the coarser scale. In particular, based on this analysis we discovered an intriguing re-entrant contraction-expansion of histone tails upon heating, which is caused by a subtle competition between enhanced ionic condensation around charged side chains [73,74,100,103,125,126] and chain entropy.

2.2 Methods

2.2.1 Simulation protocol

Since there is no reliable structural data on histone tails, we have built initial structures based on the available amino acid sequences [102]. Histone tails have not been uniquely defined in literature. For instance, histone tails are biochemically isolated by trypsination, e.g cleaving at the so-called "weak points", dividing disordered and ordered regions. However, the cleavage point does not always coincide with the residues of histone tails that are adjacent to the exit point of nucleosomal DNA [62]. We have constructed histone tails that are slightly longer than their biochemical definitions, thus following a structural viewpoint. The tail lengths are 38 residues for H3, 26 residues for H4, 23 residues for H2B and 14 residues for H2A. All the simulations were carried out using AMBER10 [43] package suite and the ff99SB protein force field [81]. Finding an appropriate forcefield for simulating IDPs can be a challenging task for several reasons. For one thing, the parameterization of

currently available forcefields is fine tuned to reproduce dynamics of proteins with single, well defined structures. Hence, a priori expectation that the default parameterization will work for IDPs is not high, and indeed extensive simulations on model unstructured peptides diagnosed serious shortcomings for many forcefields [72,154]. Luckily the comparison of explicit solvent simulations with NMR structural and relaxation data established adequacy of the ff99SB force field for IDPs [72, 136]. In a recent report it was suggested that in specific cases ff99SB might slightly underestimate the helical propensity of polyalanine model peptides [56]. Therefore, to validate the robustness of our results, we have carried out additional simulations with modified ff99SB, finding that simulations with both force fields lead to virtually identical results.

After constructing the initial structures for histone tails in a fully stretched state, we performed preliminary minimization and equilibration steps in the GBSA implicit solvent [109] to bend straight conformations to some degree in order to save computational resources associated with relaxing conformations in explicit solvent. Afterwards, each histone tail was immersed in TIP3P explicit water boxes (H3: $71 \times 61 \times 56$ $\mathring{A}^3$, H4: $56{\times}45 \times 44$ $\mathring{A}^3$, H2B: $53{\times}44 \times 42$ $\mathring{A}^3$, H2A:$48{\times}46 \times 45$ $\mathring{A}^3$) with distances between the farthest atoms of histone tails and box edges set to ~ 12 $\mathring{A}$. Particle Mesh Ewald summation technique was used for all electrostatic calculations with 12 $\mathring{A}$ real space cutoff. Periodic boundary conditions were used in all simulations.

Ions were added to neutralize uncompensated charges and further salt (NaCl) was added to represent 0.150 M ionic concentration and hence mimic the physio-

logical environment. System preparation included minimization of protein chains with the rest of the system fixed and subsequent minimization of the full system. After minimization steps, leapfrog integration with 2fs timestep in NVT ensemble was used to propagate dynamics of all atoms, where the latter were coupled to a Langeven bath with 2ps collision frequency and with weak restraining force on protein [43]. After 200ps of restrained NVT simulation, the system was equilibrated without restraints in the NVT ensemble for 300ps followed by 1.5ns density equilibration in the NPT ensemble. SHAKE [84] was used to constrain all bonds containing hydrogen atoms.

Before running replica exchange simulations each system was replicated and each replica was slowly heated and additionally equilibrated at the predefined target temperatures. Temperatures for REMD were chosen based on the criteria of a good overlap between energy distributions of neighboring replicas, guarantying significant acceptance rates [87, 141, 142]. The temperature range was chosen between 300-450 K, with 2-3K spacing resulting in 50-54 replicas and 30-35 percent expected acceptance rate estimated with the help of T-REMD server [42]. The time interval between the exchange attempts was set to 5-10 ps. Each replica was simulated at the constant temperature and volume (NVT) for 55-60 ns resulting in cumulative 3 microseconds of sampling time for each histone tail. The first 10-15 ns of all the trajectories were discarded, to allow for initial equilibration, while the rest of the trajectories were used for the subsequent analysis.

2.2.2 Principal Component Analysis (PCA)

We performed a dihedral PCA (dPCA) [47,105] on the single temperature trajectories sampled by replica exchange molecular dynamics. Since we are interested in detecting states with residual order, using dihedral PCA (dPCA) is a natural choice, because dihedral angles are the main degrees of freedom responsible for backbone flexibility and formation of secondary structural elements. In previous studies dPCA was used to scrutinize native state dynamics of small globular proteins [104, 162] and large scale conformational rearrangements of hydrated proteins [77, 120, 132]. Commonly, a few PC modes are used to reduce dimensionality of conformational space and allow investigation of dynamics on simplified landscapes [164]. For instance, the main PC mode served as a good order parameter successfully discerning states which undergo major conformational change between open and closed forms of elastin [132]. In a different work, following splitting of basins in successively higher PC dimensions allowed mapping of protein's conformational substates into a hierarchical tree [104, 162].

In dPCA, the covariance matrix $\mathbf{C}$, is constructed using sines and cosines of (ϕ,ψ) protein dihedral angles $g(t_n) = \{\sin\phi_1(t_n)\sin\psi_1(t_n), \cos\phi_2(t_n), \cos\psi_2(t_n)\ \ldots\}$, in order to avoid problems of discontinuity and multivalueness associated with angular variables [47, 105]. Each principal component is a basis vector in the high dimensional conformational space of the macromolecule, along which motions occur corresponding to the greatest variance in the data. After diagonalizing the covariance matrix (Eqs. 2.1 and 2.2) one obtains a set of orthogonal PCs and corre-

sponding eigenvalues (diagonal elements of Λ), where the former indicate how atoms are displaced in a particular mode from the time averaged structure. Thus, most of the interesting dynamics is contained in the PCs with the largest eigenvalues – these capture most of the essential macromolecular motions. After finding the PCs we used projections of the trajectory along two main PCs to map out protein's free energy surface in these collective coordinates (Eq. 2.3).

$$\mathbf{CM} = \mathbf{M}\Lambda \ , \quad C_{ij} = \langle (q_i - \langle q_i \rangle)(q_j - \langle q_j \rangle) \rangle, \tag{2.1}$$

$$v_i = \mathbf{G} m_i \ , \qquad m_i m_j = \delta_{ij}, \tag{2.2}$$

$$\Delta F(v_1, v_2) = -k_B T \ \ logP(v_1, v_2) - F_{min}, \tag{2.3}$$

Where $\mathbf{G}$ is a trajectory matrix where each column contains all the sampled values of the individual dihedral angles (time series of $g_1, g_2 \ldots$). The m_i (the i-th PC) is an i-th column of eigenvector matrix ($\mathbf{M}$) corresponding to the i-th eigenvalue(Λ_{ii}) and v_i is the projection of trajectory onto ith PC. We identified distinct basins on the constructed free energy landscape by enclosing them within squares and estimate their structural heterogeneity by computing the corresponding pairwise-q values (often used as an order parameter in spin glass physics [114]) among conformations within each square.

$$q_{AB} = \frac{1}{N_{\text{pairs}}} \sum_{i>j} \exp(-(r_{ij}^A - r_{ij}^B)^2), \tag{2.4}$$

$$q = \langle q_{AB} \rangle_{basin}, \tag{2.5}$$

where q_{AB} is the structural overlap between the A and B conformations, assuming values from 0 (no structural resemblance) to 1 (identical structures). In Eq.

2.4, N_{pairs} is the total number of C_α atom pairs and r_{ij}^A indicates the pairwise distance between C_α atoms in conformation A which are correspondingly labeled as i and j. Overall, larger value of q in any specific basin indicates potential presence of a deep trap state(s), since then conformations sampled within that basin will show high structural resemblance to each other.

2.2.3 Scaling relations

To classify the compaction state of various histone tails, we used scaling relations for globular and thermally denatured proteins to compare our computed average radius of gyration values, Rg, with the typical values for native and random coil-like states of proteins having the same number of residues:

$$Rg_{glob}(N) = 2.2N^{0.38}, \tag{2.6}$$

$$Rg_{denat}(N) = 2.02N^{0.60}. \tag{2.7}$$

The relation for globular proteins is based on a best fit of power law dependence of Rg on sequence length for a subset of proteins in PDB [137]. Similarly, for denatured proteins we use the relation that was found by the best fitting of Rg values obtained through simulation and experiments [65].

Furthermore, we compared the chain statistics of histone tails with the behavior of an ideal chain with excluded volume interactions, where we used des Cloizeaux equation [64] (Eq. 2.8) to estimate the probability density of end to end distance of a polymer chain. This equation was derived using renormalization group theory

by incorporating fast decay at large distances and low probability of end contacts into the ideal chain probability density function [64]:

$$P(X) = CX^{0.269} exp(-1.2X^{2.427}), \tag{2.8}$$

where the C is a constant determined by normalization condition ($\int 4\pi X^2 P(X)\,\mathrm{d}X = 1$) and $X = R/\langle R\rangle$. It should be noted that des Cloizeaux equation is exact only in the asymptotic limit of long and uniform chains which have simple pairwise interactions that decay with sequence wise separation.

2.3 Results and Discussion

2.3.1 "Order in Disorder": Secondary structure forming propensities of histone tails at physiological conditions.

To produce realistic conformational ensembles of histone tails we have carried out long-time REMD simulations in a wide temperature range (300-450 K), taking into account explicit solvent environment and ions. Afterwards, we employed various physically motivated means to classify and catalog the sampled conformational space. The results outlined in this section clearly demonstrate that despite possessing sequences atypical for native globular proteins, histone tails do show strong propensities for forming secondary structural elements, at specific local spots.

In our simulations we found residual secondary structural elements in three out of four histone tails. The H4 tail was enriched in beta hairpin conformations, the H3 and H2B tails had helical content and the H2A tail showed no structural

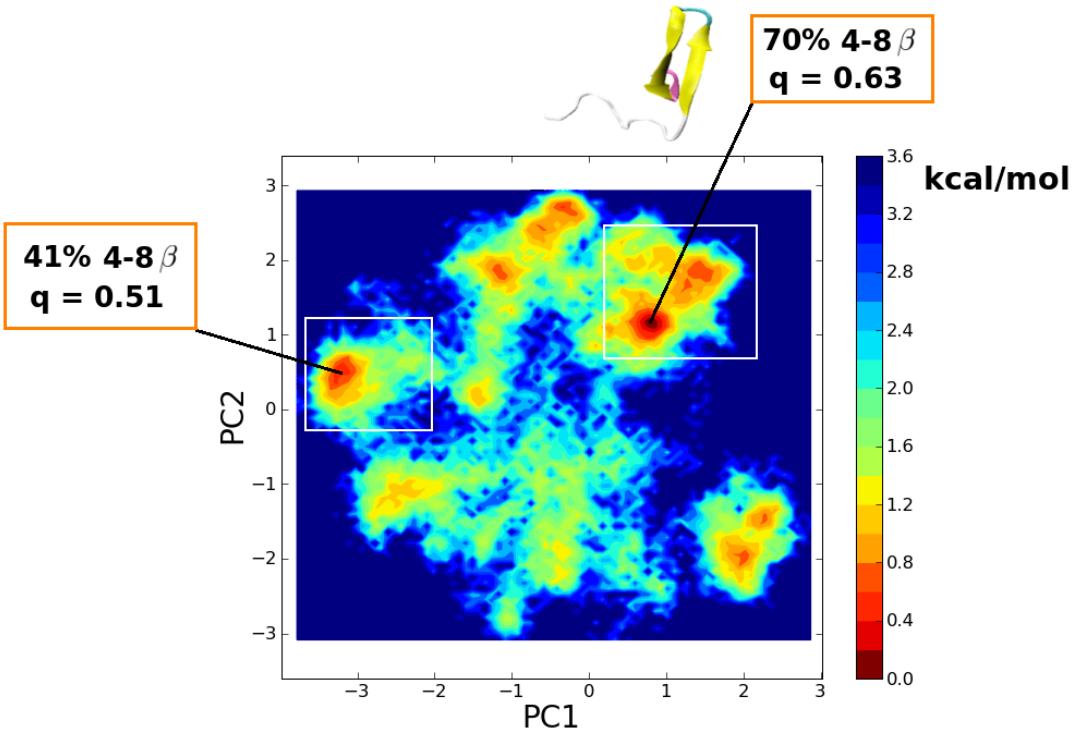

Figure 2.2: Free energy projection of the H4 tail dynamics at 300 K into its two main principal components is shown, based on Eq. 2.3.

features throughout the whole simulation (see Fig. 2.1 for representative snapshots). In the H3 chain, there are 2-3 regions that strongly favor formation of alpha helices, implying that upon binding an extensive helix formation could follow. In the H4 tail, half of the chain in the C-terminal segment (Res 12-26) forms a beta hairpin, while the N-terminal segment (Res 1-12) remains fully disordered. The H2B histone tail showed weaker propensity to form secondary structures. To disentangle interesting conformational modes from pool of states sampled by single temperature REMD trajectories we applied PCA in space of (ϕ, ψ) peptide backbone torsional angles (dPCA). Two-dimensional free energy surfaces for all four histone tails were mapped using first two PC modes corresponding to the largest eigenvalues, since these modes

move on the slowest timescales.

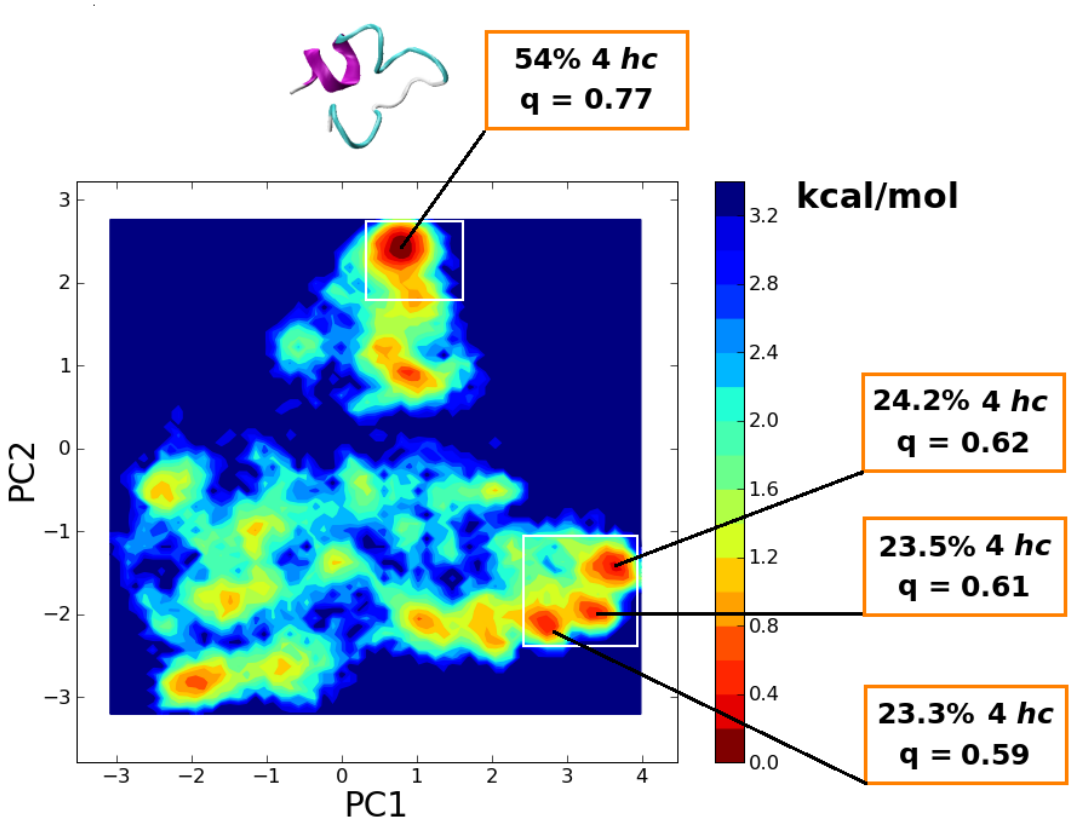

Figure 2.3: Free energy projection of the H2B tail dynamics at 300 K into its two main principal components is shown, based on Eq. 2.3.

Obtained two-dimensional free energy landscapes (Fig. 2.2- 2.4) demonstrate that conformational space of three histone tails (H4, H3 and H2B) are well defined by a handful number of distinct basins. In each basin (see the white squares on Fig. 2.2- 2.4), we have computed the percentage of conformations that possess residual secondary structure, finding a noticeable variation in the degree of secondary structural content among different basins. To quantify the conformational heterogeneity inside basins, we have also computed the average mutual structural overlap q between all conformations within each basin. When a conformational trap(s) with

deep free energy is present in a given basin, it tends to attract many visits during peptide dynamics, resulting in similarity of many conformational snapshots to each other, hence, high q values. Our subsequent analysis indicated that only a weak correlation exists between residual secondary structural content and basin's q, which is due to the fact that significant portions of all chains are in disordered conformations. Here we should also emphasize that the proximal basins on the PC free energy surface are kinetically accessible [48], e.g. the states in neighboring basins are structurally closer to each other compared to states in more distant basins.

Landscape topographies of the H4, H3 and H2B tails show a signature of well defined multiple basins of various depths, whereas the basins are more shallow for the H2A tail. We have indicated the thermodynamically dominant basins with white boxes in Figures 2.2 and 2.3. On average, the basin depth difference between dominant and other basins is $\sim 2 - 3k_BT$, hence we expect that the dynamics will mostly be determined by moderate lingering in the dominant basins and frequent transitions among dominant and other basins. When examining the specific features of each tails' free energy landscapes, we found partially structured conformations along with fully disordered ones in all basins of the H4, H3 and H2B tails, where in all cases the dominant basin had the highest content of partially structured states (see Figs 2.2- 2.4). The structural overlap parameter, q, was also highest in these basins, indicating the presence of thermodynamically dominant, specific conformations. In particular, the landscapes of the H2B tail (Fig.2.3) and the H3 tail (Appendix) contain basins with varying content of alpha helical segments, while the H4 tail basins were enriched in beta hairpins (Fig. 2.2). On the other hand,

the free energy landscape of the H2A tail (Fig.2.4) shows a more connected web of basins with the absence of conformations possessing any secondary structure, which is primarily attributed to the fact of H2A tail having the shortest sequence with the highest charge per residue (Fig.2.1). The broad and connected basins on the landscape of the H2A tail (see Fig. 2.4) correspond to random coil like states and the deepest basin consists of mostly stretched conformations with relative R_g values significantly exceeding R_g of other histone tails, as discussed below. Our present results are consistent with prior experiments [55] and secondary structure prediction results [79], which support the view of the H2A tail being a random coil and the other tails possessing residual structure.

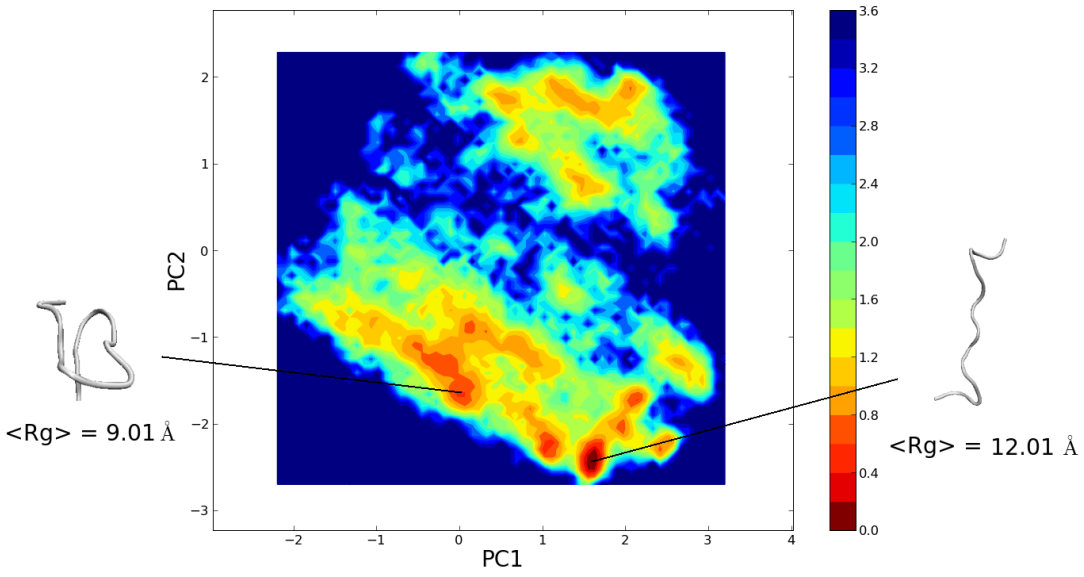

Figure 2.4: Free energy projection of the H2A tail dynamics at 300 K into its two main principal components is shown, based on Eq. 2.3.

Next, we closely examined the distributions of (ϕ, ψ) dihedral angles for several key residues in various histone tails. In the H4 tail, for instance, we found

an interesting structural feature that adds an extra layer of stability to beta hairpin state relative to its disordered conformation. There are two pairs of LYS-ARG residues in a beta turn that reduce the conformational flexibility of the turn, restricting it to a small number of states. The resultant states favor beta hairpin over the other conformations, where if the latter were realized that would have inevitably led to a steric clash between positively charged side chains of LYS and ARG. On the contrary, in globular proteins GLY is the residue that frequently resides in the beta turns making them more flexible and thereby facilitating formation of compact secondary structures.

The LYS-16 residue, which in our simulations is frequently locked into a specific rotameric state in the beta turn, is a well known hot spot for post-translational covalent modifications [63, 66]. Hence, based on the crucial role that LYS-16 plays in stabilizing H4 tail's beta hairpin, observed in the current work, we speculate that covalent modifications may potentially shift the H4 tail conformational equilibrium, providing an additional mechanism for signaling through post-translational modifications. Additionally, in vitro experiments have shown that homogeneous mono-acetylation of the H4 tail undermines the stability of chromatin fiber to an extent that removal of whole H4 tail is equal in its effect [135]. Our current results suggest a possible molecular level explanation, namely that acetylation of H4 tail triggers partial or full disruption of beta hairpin, leading to modulation of binding affinity towards linker DNA. Ramachandran plots of several other key H4 residues showed that LYS and ARG residues sample diverse conformational states in the disordered N-terminal region, while the analogous residues in the C-terminal re-

gion are highly constrained (see Fig. 2.5a). This conformational dissimilarity of two halves of the H4 chain was further explored by computing radii of gyration of two terminal segments. Obtained distributions of Rg values demonstrated that the part with enhanced secondary structural content (C-terminal) on average has less variability in its size distribution (Fig. 2.5b), thus supporting the global view of H4 tail as "half-ordered, half-disordered". In the H3 tail, there are 3 LYS-ARG regions which, as in the H4 tail, form beta turns aiding in overall compaction. However, contrary to the H4 tail, local sequences of the H3 tail around beta turns favor alpha helices, producing an overall enrichment in alpha-helical transient states. In the H2A tail, all residues were unconstrained and spanned almost all allowed regions on the Ramachandran diagram.

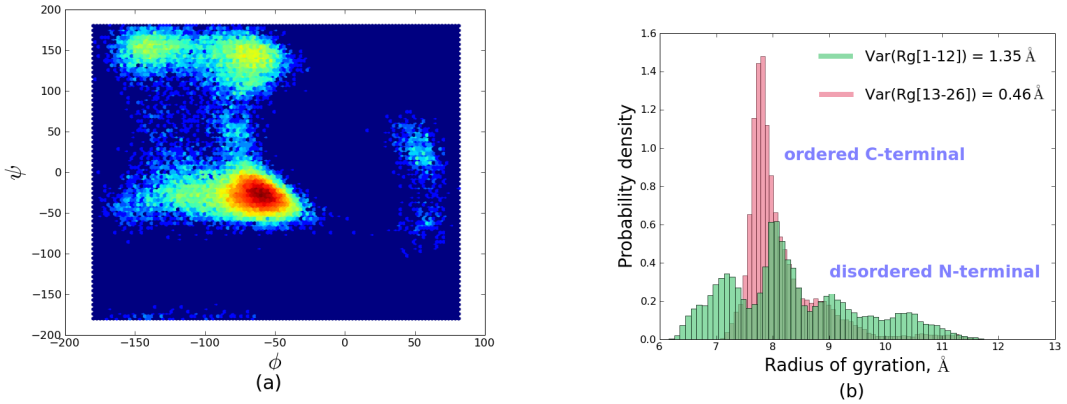

Figure 2.5: (a) Ramachandran plot of Lys-16 in the H4 tail is shown (b) Distributions of radii of gyration for the N (residues 1-12) and C (residues 13-26) terminals of the H4 tail are plotted.

Next, we explore the size distributions of histone tails. We computed distributions of radii of gyration for the conformations sampled from the equilibrium ensem-

bles of histone tails and used mean Rg values for comparison with well investigated classes of globular and thermally denatured proteins (Fig.2.6). From this comparison one immediate finds that all histone tails except H2A, on average have sizes that are closer to native globular states than random coils. In particular, despite their disordered nature, conformational ensembles of H4/H3/H2B histone tails significantly deviate from random coil statistics, and instead are better described as collections of relatively compact "molten globular" (MG) states (see Fig. 2.7). This is shown by the mean values of their radii of gyration, which for the H4 tail differs by $\sim$30 % from the corresponding hypothetical native state (see equations 2.6 and 2.7) of a globular protein with the same length ($\langle Rg_{H4} \rangle = 10.4\mathring{A}$, $Rg_{glob} = 7.6\mathring{A}$, $Rg_{denat} = 14.3\mathring{A}$). The H3 tail's degree of compaction ($\langle Rg_{H3} \rangle = 11.6\mathring{A}$, $Rg_{glob} = 8.8\mathring{A}$, $Rg_{denat} = 17.91\mathring{A}$) is qualitatively similar to the H4 tail, though the spread of Rg distribution is much higher. The H2B tail, which is only 2 residue shorter than H4, showed higher degree of compaction, with $\langle Rg_{H2B} \rangle$ being only $\sim$ 20 % higher then the corresponding native globular form ($\langle Rg_{H2B} \rangle = 9.12\mathring{A}, Rg_{glob} = 7.36\mathring{A}, Rg_{denat} = 13.59\mathring{A}$). This enhanced compactness compared with the H4 tail may be rationalized by the lower ratio of H2B tail's net charge to its sequence length (H2B: 0.25, H4: 0.35) and also a higher ratio of hydrophobic structure forming residues. For the H2A tail, the situation is qualitatively different as indicated by its Rg distribution (Fig.2.6b), which puts the H2A conformational ensemble closer to random coil like states ($\langle Rg_{H2A} \rangle = 9.37\mathring{A}, Rg_{glob} = 6.0\mathring{A}, Rg_{denat} = 9.86\mathring{A}$).

From a polymer physics perspective, histone tails are polyelectrolytes with intricate conformational behavior stemming from propensity to form secondary struc-

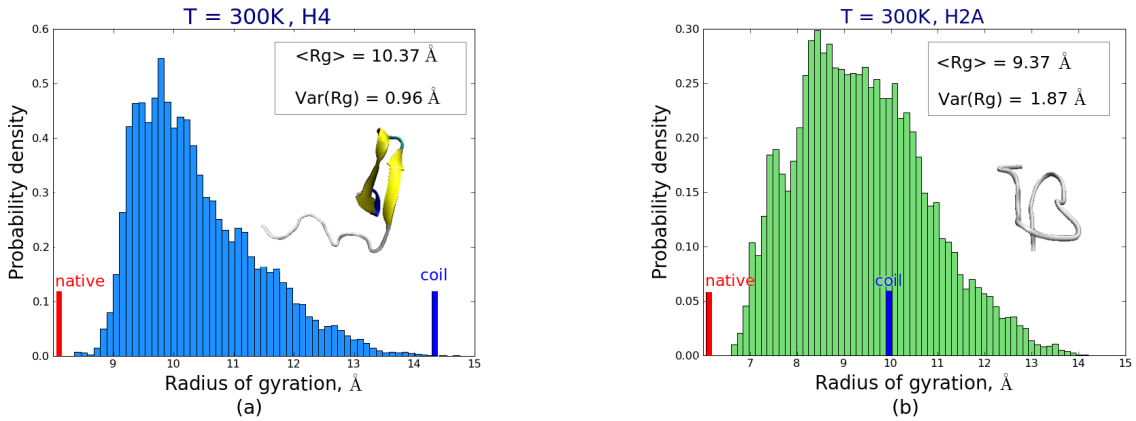

Figure 2.6: Distributions of radii of gyration for the (a) H4 and (b) H2A tails are shown at 300K. Red and blue bars indicate expected sizes for a globular folded protein and a random coil respectively, for a hypothetical peptide of the same length.

tures, having nonuniform charge distribution and hydrogen bonding patterns. Because of the presence of such interactions, one would expect a significant deviation from classical polymer theories that are mainly based on simple coarse grained pairwise potentials between uniformly shaped monomers. Hence, we also computed the end-to-end distance distributions for all histone tails, and compared them with the corresponding distributions obtained from Gaussian chain in the excluded volume limit as described by des Cloizeaux equation (Eq. 2.8). Indeed, from comparisons with des Cloizeaux curve (Fig.2.8), we see a drastic difference with the low temperature end-to-end distance distributions of the H4, H3, and H2B tails (data for H3 and H2B not shown). The multi-peak nature of these distributions is most likely caused by the presence of secondary structural elements, where the latter introduce additional complexity in monomer-monomer interactions that is not captured by des Cloizeaux equation. For instance, in the H4 chain there is a significant popu-

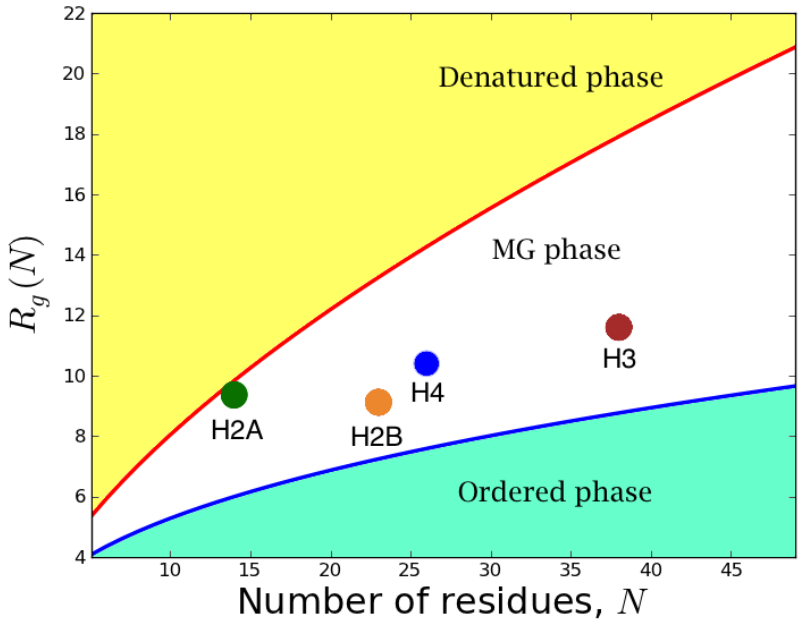

Figure 2.7: Approximate protein phase diagram, showing denatured, molten globular and native globular regions.

lation of both states with on average close approach of monomers and states with higher degree of separation (Fig.2.8a). These results clearly point to the limitations of classical polymer physics ideas for characterizing even disordered protein chains, which in general demand inclusion of more structural details. On the contrary, the H2A tail is reasonably well approximated by des Cloizeaux expression (Fig.2.8b), indicating that it is a random coil. The peak of distribution is shifted relative to des Cloizeaux curve due to the presence of multiple charges. Overall, our results suggest that worm-like chains models are not expected to adequately capture conformational dynamics of histone tails.

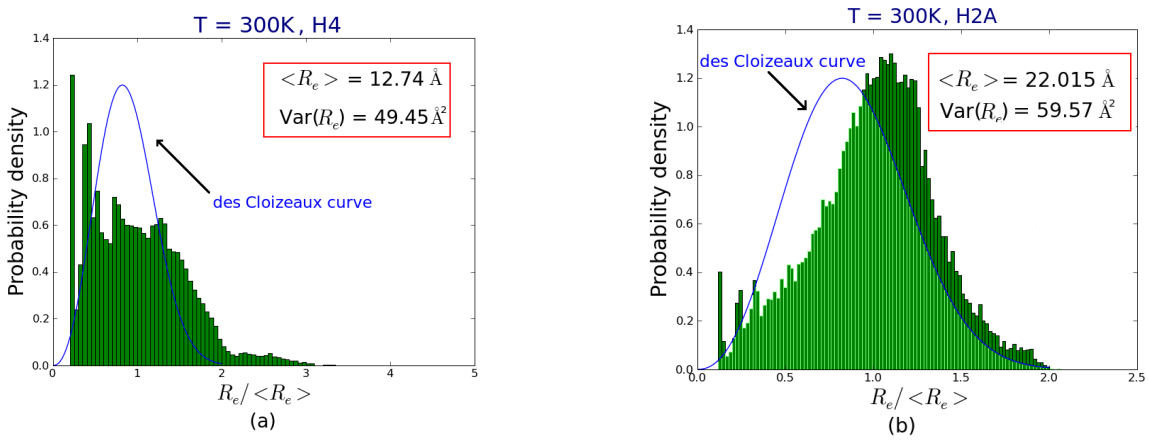

Figure 2.8: End-to-end distance histograms of the (a) H4 and (b) H2A tails are shown at T=300K and compared with the des Cloizeaux equation predictions.

2.3.2 Thermal Denaturation and Ionic Effects in Histone Tails.

Histone tail sequences are unusual because they contain a plethora of positively charged residues. This aspect is not surprising since histone tails regulate chromatin structure and dynamics by interacting with negatively charged DNA and acidic patches of histones cores. Hence, electrostatics plays a key role in their functionality. In particular, the distribution of counterions around charged surfaces of histone tails is expected to play crucial role in interactions of tails with DNA, through release and re-association of ions around biomolecular surfaces, as well as influencing the equilibrium ensemble of histone tail conformations [74, 140]. To gain further insight into the histone tail electrostatics, we investigated the nature of ionic environment around protein backbones, including temperature dependencies. One remarkable aspect of polyelectrolytes is their ability to undergo rapid contraction into compact forms upon increase of the counterion concentration [6, 78, 94]. The main driving force for this transition is screening of the electrostatic repulsion among the charges

of the chain, which, in turn, diminishes swelling of the conformations that occurs due to these repulsive interactions. In our case, however, concentration of the salt is held constant and the parameter whose change induces chain contraction is the temperature. This temperature induced chain collapse has not been studied in prior theoretical works on flexible polyelectrolytes [6,45,59,60,78,82,96], and here we give qualitative explanation about the cause and explain the specifics of the transition for different histone tails.

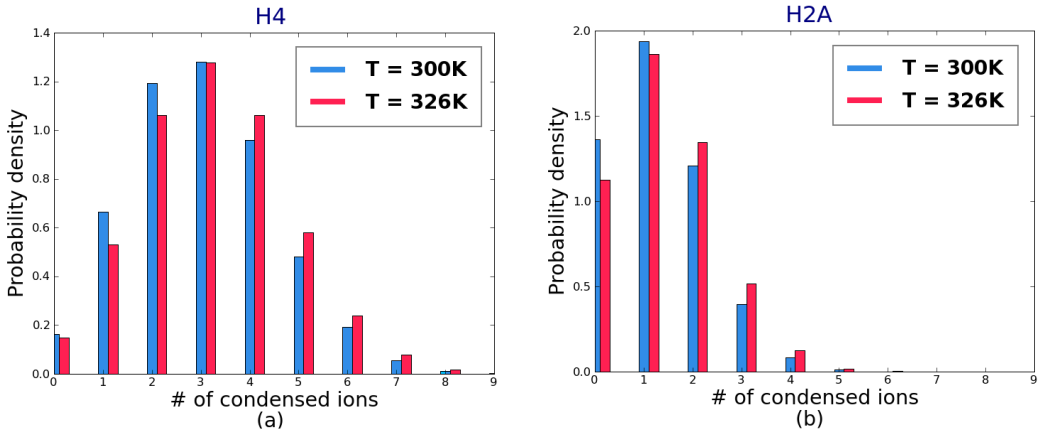

Figure 2.9: Temperature dependencies of ionic condensation around backbones of the (a) H4 and (b) H2A histone tails are shown at 300 K and 326 K.

First, to characterize the low resolution details of ionic association around histone tails, we computed the total number of anions (Cl^-) around positively charged backbone using as a threshold a distance between nitrogens of ARG and LYS and chloride ions being less than or equal to the Bjerrum length ($\sim$7.5 Å) at ambient temperature. Interestingly, upon heating we observed consistently increasing accumulation of ions around both histone tails (2.9). This type of increase of ionic condensation with temperature was already observed in a number of MD simulations

of other proteins [73, 74] and in some cases was shown to be sequence specific [100]. One way to understand the observed elevation of ionic adsorption at high temperatures, is to take into account the temperature dependence of the dielectric constant. The experimentally found expression for the dielectric constant of water [121] is : $\epsilon(T) = (\frac{T^*}{T})^{1.4} \sim T^{-1.4}$, hence the Bjerrum length ($l_B = \frac{e^2}{\epsilon(T)k_BT}$) of an aqueous medium is effectively a weakly increasing function of temperature: $l_B \sim T^{0.4}$. Therefore, qualitatively, at high temperatures ionic interactions are stronger, which under favorable conditions can lead to enhanced ionic condensation.

Furthermore, we observed consistently greater accumulation of Cl^- ions around residues in disordered N terminal regions compared to ordered C terminal regions. This trend seems to arise because of larger steric accessibility of disordered regions for mobile ions. Accumulation of ions around chains increases the ionic screening, resulting in diminution of electrostatic repulsion among charged side-chain groups. This enhanced screening in turn explains why histone chains instead of expanding with moderate increase of temperature, somewhat contract. For the H4/H3/H2B tails, the effect of ionic condensation and subsequent chain contraction were more pronounced than for the H2A tail. This difference in ion adsorption at elevated temperatures is caused by the disruption of secondary structural elements in the H4/H3/H2B tails which ultimately leads to more surface exposed conformations, with charged side-chains having fuller access to ions floating around. Hence, at all temperatures above T=300K the H2A tail on average attracts less mobile ions per chain charge compared with the other tails.

The main principal modes (PC 1-5) at moderately high temperatures (T ∼320-

326K) showed the emergence of new compact states, as evidenced by the appearance of additional deep minima (see Appendix), which vanish upon further heating. Additionally at temperatures corresponding to collapse states, we have found sharper decay of eigenvalues (e.g., Λ_{ii} from Eq. 2.1) when arranged in decreasing order (see Appendix). This interesting observation is explained by the idea that in collapsed chains dihedral angles are fluctuating in a more restricted range of values and motions are relatively damped, because of tighter monomer packing. These restrictions naturally lead to a greater correlation between various dihedral angles, which as a result increases the contribution of first two PC's into total atomistic motions. On the other hand, when the chain is disordered and samples variety of weekly correlated conformations, dihedral angles make frequent transitions between allowable Ramachandran regions and cause increased PC degeneracy. For instance, first 2 PCs were able to capture ~ 50 % of motions for a globular native protein eglin c [104], wheres for IDPs like histone tails the contribution of two main PCs are below 20% (see Appendix).

Finally, the Gaussian chain model with volume interactions seems to be more adequate for histone tails at very high temperatures (see Appendix), which can be explained by both unraveling of secondary structure elements, as well as better screening of histone tail charges by mobile ions, leading to increased relative contribution of excluded volume interactions. In a notable contrast with regular globular proteins, which in general expand abruptly in an "all or none" fashion near the melting temperature, histone tails first contract when temperature is increased, then expand at higher temperatures. The dPCA free energy landscapes clearly show

how upon further heating, these collapsed states gradually and slowly swell, with fragmented PC basins increasing in number and coalescing into a large, smooth single basin at high temperatures (Appendix). This collapse/expansion re-entrant transition results from the competition between the electrostatic interactions mediated by mobile ions, discussed above, and chain entropy, which eventually wins at high enough temperatures.

2.4 Concluding Remarks

Our analysis of 3 microsecond long REMD simulation trajectories showed that equilibrium conformational ensembles of histone tails are comprised of states with various degrees of residual order. We further characterized the ensemble behaviors by mapping out simplified free landscapes using the first two PCs. On every landscape we identified highly populated states and classified them by using secondary structural content and average pairwise conformational overlap values. We found that the H4 tail is enriched in beta hairpins, H3 and H2B tails contain flickering alpha-helical segments, while the H2A tail is disordered, with partially stretched conformations.

Based on our findings we propose that states with high content of secondary structural order may be important for binding to linker DNA and acidic patches since rigidity of these elements provides higher affinity and lower entropic penalty for binding. From comparisons of chain size distributions we conclude that H4, H3, and H2B tails show behavior consistent with compact molten (Fig. 2.7) globular

states such as appearing on protein folding pathways, while the H2A tail samples predominantly random coil like states.

In summary, our analysis demonstrated that the presence of secondary structural elements in histone tails creates a population of conformations that are drastically different from the random coil like states described by Gaussian chain with excluded volume interactions or worm like chain analogues. These results invalidate the naive picture of histone tails which are viewed as a collection of mostly disordered and structurally uncorrelated states. In addition, our investigation of histone tails' ionic environment and their temperature dependence revealed the remarkable nature of denaturation of histone tails which occurs via re-entrant mechanism, when the chain at first contracts and then slowly swells as temperature is increased. Finally, in the H4 and H3 tails, we found that a specific sequence motive, LYS-ARG, locally favors beta turn formation and kinks the chain making it more compact and favorable for secondary structure formation. Post-translation modifications of the corresponding LYS residues may potentially unlock these local structural motifs inducing large scale conformational changes, leading to drastic changes in tail-DNA and tail-histone core binding free energies, which can potentially account for the corresponding disruption of chromatin fiber observed in *in vitro* experiments. The main results of this chapter should be helpful in interpreting future experiments on histone tails and better understanding of their functional roles.

Chapter 3

Molecular switch mechanism for the H4 acetylation induced control of gene expression.

The chapter is based on the submitted work of the author:

D. A. Potoyan, G. A. Papoian; (2012)

3.1 Introduction

Regulation of cellular processes by the post-translational covalent modifications of the histone proteins is a route that is widely exploited by all the eukaryotic cells [5]. While the cores of histones are essential for packaging the genomic DNA inside the cell nucleus, which is further mediated by the terminal histone tails, the highly specific post-translational modifications of both histone core and tail residues allows the cell to achieve a broad control of the accessibility of the genetic information. Hence, not surprisingly, the DNA sequence alone does not determine the ultimate fate of the gene expression. In particular, histone tail modifications provide one more layer on top of the DNA sequence that guards the canonical pathway from DNA sequence information to biological action. In this way, environmental factors can contribute to the diversification of the phenotype via epigenetic changes. In his review, Schiessel compared the well-organized chromatin to a library, where books

(genes) are arranged on the shelves for an easy access [131]. To continue the analogy, the histone tails may then be regarded as the librarians who control when and which books are checked out or returned. Thus, while the DNA contains static information about each proteins composition, the dynamics of gene expression are, in significant part, determined by a variety of proteins that covalently modify histone tails, hence triggering specific biochemical reactions and structural rearrangements. The following modifications of histone tails have been observed among others [5]: methylation, acetylation, ubiquitinations, and sumoylation. The different combinations of these are recognized as a code for activating or suppressing a particular biochemical event [139, 173, 176]. There are two ways that covalent modifications can modulate the timing of the gene expression [176]: one is by recruiting specific binding agents to the modified sites, and the other is by direct physical changes in histone tail-nucleosome interactions. Our present work deals with the acetylation of the H4 tail, the mode of action of which may belong to the second category [191] Some of the end effects of histone tail acetylation are now well known and documented [131, 168, 170, 176]. However, there seems to be little if any molecular level rationale about how the acetylation of the H4 tail can induce the observed dramatic changes in the chromatin organization.

In our recent work, we carried out a systematic study of the conformational preferences of histone tails, focusing mostly on the physical aspects of the individual tails, hence, discussing only briefly the possible biological implications of tails physical interactions and polymer chain statistics. In particular, we found that some of the tails are not fully disordered, but their energy landscapes are organized in

special ways [184]. This work continues and complements our previous endeavor, where we now focus more on the functional aspects of histone tails, choosing the important H4 tail as a case study. We use an explicit solvent all atom simulations and simple concepts acquired from the spin glass theory to try to shed some light on one of the fundamental questions of the chromatin science: why and how do covalent modifications influence the conformational and binding propensities of the histone tails?

In the present study, we focus on an important and well-studied modification of H4 tail, namely the acetylation of LYS-16. This particular modification is well known for being implicated in the transcriptional activation [177,186]. From in vivo experiments, it was found that acetylation weakens the chromatin packing, allowing transcriptional factors to access the specific gene sequences [186]. In another landmark in vitro experiment with reconstituted nucleosomal assays, the homogeneous mono-acetylation of H4 tails at the LYS-16 residue lead to the massive disruption of the dense 30 nm chromatin fibers [135]. Based on these results, one can argue that reduction of charge is not enough to explain the effect of acetylation, since raising the salt concentration was found to not reverse its disruptive effect [80]. Therefore, we hypothesize that acetylation mediates the destabilizing effect of the acetylated H4 tail through a physical mechanism which is more elaborate than what the simple electrostatics arguments would suggest. Unfortunately, not much is known about the forces that create the mentioned bulges in the poly-nucleosomal arrays, which weaken the packing of chromatin fibers.

3.2 Results and Discussion

To address some of the issues raised, we generated the conformational ensemble for LYS-16 acetylated H4, by performing 3 microseconds long replica exchange MD simulation (REMD) following the same protocol as outlined in the previous work [184]. Our prior study of unmodified histone tails revealed that LYS-16, which resides in the sterically crowded beta turn, might destabilize the beta hairpins. In the present study, we find that acetylation of LYS-16 indeed reduces the beta hairpin content in the equilibrium conformational ensemble. We employed the distribution of the pairwise structural overlap values, q, as a means to study the effect of acetylation on conformational preferences of the H4 tail. The parameter $q_{ij} \sim \sum_{a,b} \exp\left[-\frac{(r^i_{ab} - r^j_{ab})^2}{2\sigma^2}\right]$ originally introduced for the study of spin glasses [114], has subsequently found wide application in the folding studies of the native globular proteins. It quantifies the structural similarity of conformations i and j on a $(q_{min} = 0, q_{max} = 1)$ scale, where higher values mean higher structural resemblance and vice versa. In the protein folding studies one has a well-defined native state and pairwise q values (by convention referred as Q values) are helpful for quantifying the departure or approach towards it. However, for the intrinsically disordered proteins, there is no such unique state. Instead one often seeks some structural patterns in the maze of conformations that these proteins sample under physiological conditions. For that purpose the distribution P(q) of the pairwise q values between all distinct conformations would be a natural extension for characterizing the conformational ensembles of the intrinsically discorded proteins.

Physically, P(q) characterizes the structural heterogeneity of the conformational ensemble. Another commonly used measure, the distribution of the radius of gyration, $R_g \sim \sum_{a,b}(r_a - r_b)^2$, is less informative in terms of revealing the emergence of new structural clustering, especially for the short chains. The comparison of the P(q) distribution for the wild type (WT) and the acetylated H4 tails immediately reveals (Fig. 3.1) that the latter contains a subset of compact conformations, which are absent in the conformational ensemble of the WT H4 tail. The chain compaction is not surprising, since acetylation reduces the positive charge of the chain, and hence, self-repulsion. It also adds to the hydrophobic interactions within the chain [94]. Hence, the introduction of acetyl group not only decreases the beta hairpin content in the conformational ensemble but also promotes the formation of more compact globules (see the inset of Fig. 3.1) with higher helical propensity (10% of conformations). Nevertheless, the beta hairpin content in the acetylated form, although diminished from the WT, still remains significant in comparison (Fig. 3.1). Overall, the qualitative trend of the conformational changes upon acetylation of LYS-16, observed from simulations, is consistent with the prior CD experiments, which reported steady increase in the alpha helical content of the H4 tails as a function of added acetyl groups [55, 153]. Explicit solvent simulations of the fully acetylated H4 tail likewise showed the robustness of the alpha helical conformations [170]. However, we note that many commonly used force fields might not reflect the quantitatively accurate ratio of the relative populations of alpha helical versus beta hairpin conformations for the disordered polypeptide chains [72, 154].

Next, we proceed to probe the impact of the H4 tail acetylation on its DNA

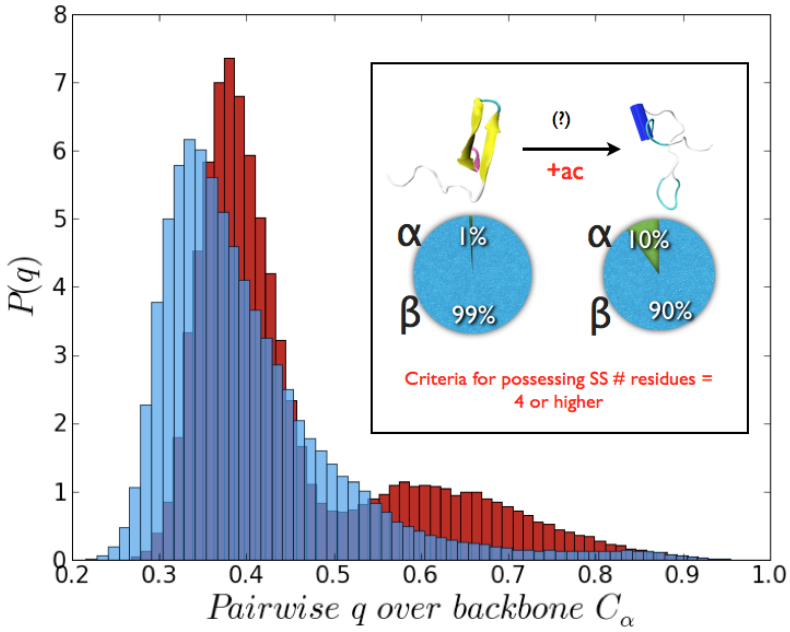

Figure 3.1: Comparison of P(q) distributions of WT(blue) and covalently modified H4 tail(light green). The small peak on the left side shows the emergence of compact states in the acetylated H4 tail conformational ensemble. The inset shows the secondary structure contents in the conformational ensembles of the isolated WT (left) and covalently modified (right) H4 histone tails.

binding propensity. We randomly chose a 20 bp DNA sequence as a mimic of some segment of the nucleosomal DNA. First, we did a control simulation of the H4 tail with the DNA for 200 ns to ascertain that our subsequent umbrella simulations with restrained DNA protein distance are not biased towards unrealistic conformations and to dissect the effect of covalent modification on the conformational changes induced by the presence of DNA. After binding to the DNA, the WT H4 tail undergoes very little structural fluctuation, which is expected for a highly charged polymer (see Fig. 3.2). The residues that are in contact with the DNA are

all part of the transiently flickering beta hairpin, which also includes the LYS-16. In the chromatin, the C terminal of H4 tails are structurally in close proximity to the DNA ramp of the nucleosome. Hence, it is a natural choice to use the center of mass distance between the target DNA base pair and the C terminal of H4 tail. Afterwards, we performed umbrella sampling along the DNA-protein distance coordinate with the spring constants that guarantee good overlap between sampling coordinate distributions.

The technical details regarding the system preparation and the MD simulation of the DNA-histone tail systems are elaborated in the Supporting Information section. Briefly, the protocol for the potential of mean force (PMF) calculation is as follows. At first, the 20 bp long DNA was positioned at 25 Angstroms from the center of mass of the H4 tail. The refined AMBER22 Parmbsc0 force field for nucleic acids [183] and the TIP3P model for the water [174], as well as the recently developed force field [175] for alkali and halide monovalent ions, were used for setting up the system and for the subsequent minimizations and all atom MD runs. We used the Weighted Histogram Analysis Method [89] (WHAM) for obtaining the DNA-H4-tail PMFs. We have performed a control PMF calculations starting from different initial conditions, which repeated the obtained results (see the Supplementary Information).

From the PMF plot of WT H4-DNA (Fig. 3.3), we find a $\sim 2k_BT$ (1.2 kcal/mol) gain in the binding free energy, indicating that thermal fluctuations can frequently and quickly break the DNA-tail interaction. This would allow the WT H4 tail to be flexible at reasonably fast timescales and easily explore extended chain

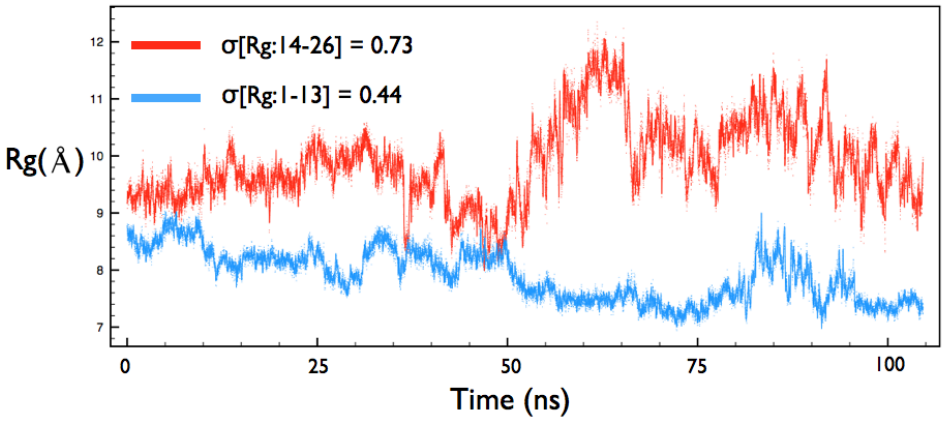

Figure 3.2: The last 100ns of the trajectory showing the structural fluctuations of in the structured (blue, residues: 1-13) and disordered (red, residues: 14-26) regions of the DNA bound WT H4 histone tail. The standard deviations of the Rg values are reduced by ∼40% when compared to the free form of the WT H4 histone tail.

conformations needed for reaching linker DNA and other nucleosomes. On the other hand, we estimate the free energy gain to be about 5 kT (3.0 kcal/mol) for the acetylated H4, which, in turn, is high enough to keep the tail glued on the surface of the DNA. The fact that the minima on the PMF of the acetylated H4-DNA is shifted to the right compared to the WT H4-DNA is simply because, throughout the whole simulation time the acetylated form remains in the dense globular-like state, while WT H4 has half of its chain protruding radially with respect to the DNAs helical axis. For the same reason, the considerable difference in the PMF depths between the WT and acetylated forms may be rationalized by noting that the highly charged flexible N terminal portion of the WT H4 tail, which contains half of its total charge, is not bound to the DNA, whereas the acetylated H4 is fully collapsed on the DNAs surface (see Fig. 3.4). Thus, despite the fact that acetyl group is reducing overall

positive charge on the H4 tail, the net binding free gain clearly favors the acetylated from of the H4. The similar binding trend previously has been reported [179] for the disordered C- terminal of the p53 protein, which upon the post-translational acetylation shows dramatic increase in its DNA binding affinity. Qualitative understanding of such behavior can be obtained by consideration of the forces that drive the protein-DNA binding. From the studies of the DNA protein complexes it has been recognized that the net binding free energy gain originates primarily from three distinct contributions [169, 189]: electrostatic interactions, hydrogen binding and shielding of hydrophobic residues from the aqueous environment. The electrostatic component alone tends to create non-specific DNA-protein complexes, which are diffuse associations [185], utilized for instance by facilitating the rapid scanning of the genome for the target sites [172]. On the other hand, the specific association with the DNA is obtained by forming extensive network of hydrogen bonds, which is further assisted by the hydrophobically driven clustering of nonpolar residues near the binding surface. Furthermore, the electrostatic calculations [180, 181] and thermodynamic analysis of numerous DNA-protein complexes [187] shows that the major driving force in fact has little to do with the non-specific polyelectrolyte interactions and is primarily accounted by the hydrophobic interactions and hydrogen bonding.

Therefore we conclude that the enhanced binding affinity of the acetylated H4 tail comes from its collapsed and more hydrophobic nature that makes more contacts with the surface of the DNA. The strong interaction of acetylated H4 with the DNA is expected to have a noticeable structural impact on the local inter nucleosomal

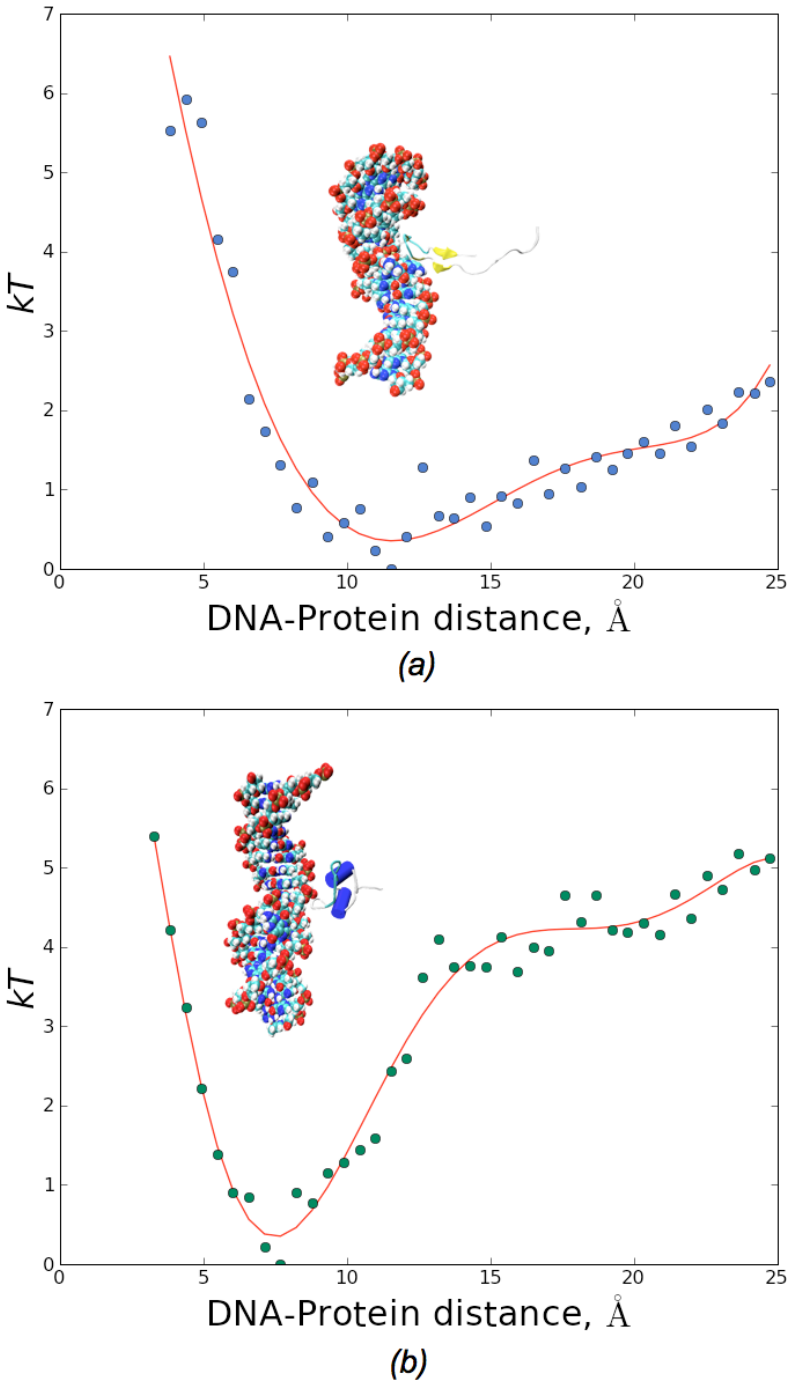

Figure 3.3: Comparison of PMF profiles of the DNA binding for the (a) WT and (b) the LYS-16 acetylated H4 histone tail.

configuration, which would then propagate down further by ultimately changing the chromatin architecture on a much larger scales [135]. A direct support for the latter

claim comes from the FRET measurements of DNA labeled nucleosomes [171, 188], which reveal distinct structural changes on the mono-nucleosomal level that are driven solely by H4 acetylation. Furthermore, these FRET experiments indicate that H4 acetylation results in tightening of the free DNA ends coming out of the nucleosome [171, 188]. This result is consistent with the idea of stronger binding of acetylated H4 tail to its own nucleosomal DNA, which is the main suggestion of the current work. In terms of the secondary structural propensities, for the WT H4 tail, we see no qualitative changes upon binding to the DNA. The H4 beta hairpin conformations are simply retained at all distances from the surface of the DNA. On the other hand, the secondary structure analysis reveals, that for the acetylated H4, the beta hairpin content is nearly zero once the tail is in sufficient proximity to the DNA surface. Meanwhile, the alpha helical content is significantly enhanced when bound to the DNA. Therefore, while for an isolated H4 tail, acetylation slightly tilts the balance of beta hairpin towards becoming an alpha helix, binding to the DNA completely solidifies the trends, resulting in full switching to an alpha-helical local conformation.

The unique role of the H4 histone tail in maintaining the chromatin architectural organization has been established by sedimentation experiments [170], which showed that, from all the tails, only the absence of H4, or equivalently its acetylation, can cause an irreversible disruption of the chromatin fiber. To the present day, the molecular mechanisms by which the H4 histone tails contribute to the stability of organized chromatin have not been elucidated. This is in part due to the fact that histone tails operate in the complex environments where they potentially in-

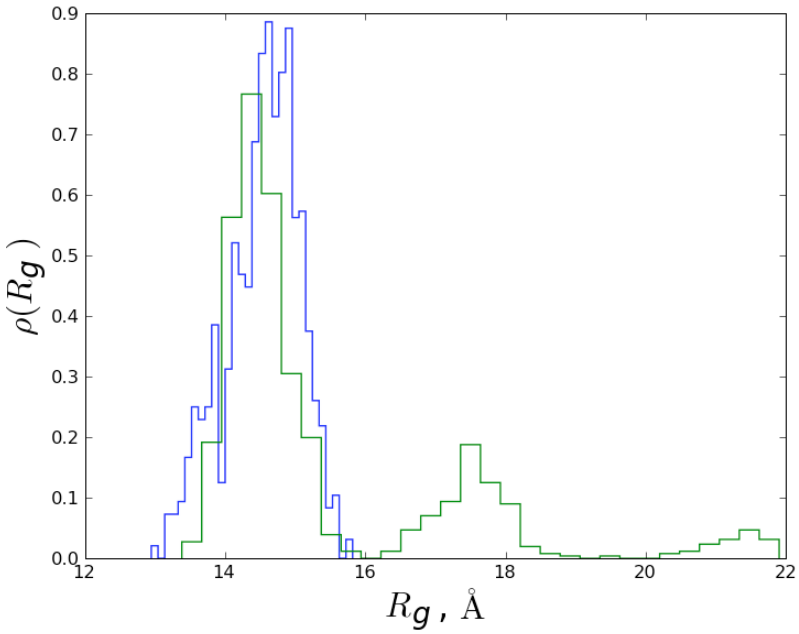

Figure 3.4: Comparison of the distributions of the radius of gyration for the bound states of the WT (green) and the acetylated (blue) forms of the H4 tail.

teract with the histone core, the nucleosomal DNA and also with the other histone tails. Among the possible clues is the finding [167] that all of the WT core histone tail domains have well-pronounced preference to bind the linker DNA - the short segment that joins two adjacent nucleosomes [103, 188]. Coarse grained molecular simulations also seem to stress the importance of the so called tail bridging effect in increasing the attraction between the nucleosome cores [182].

3.3 Concluding Remarks

In the context of the chromatin folding our results suggest a simple molecular level mechanism that explains how the acetylation of the H4 tail can cause the local unraveling of the chromatin fibers, which is necessary for allowing the access

for the various binding agents. In a nutshell, we discovered that the acetylation induces partial chain collapse and results in the tighter binding of the H4 tail to the DNA, including, potentially to its own nucleosomal DNA. This, in turn, would significantly suppress the access of H4 tails N terminal segment to the regions outside of the nucleosome, and hence, disallow the tail to mediate stabilizing interactions with the neighboring linker DNA and nucleosomes, leading to the formation of local bulges in the dense poly-nucleosomal arrays. Hence, we suggest that WT H4 tail can freely explore its conformational space, while acetylation is an auto-inhibitory molecular switch which leads to H4 tails sequestration via enhanced binding to its own nucleosomal DNA.

Chapter 4

Estimating the free energy difference of allosteric transitions: Adenylate Kinase as a test case

The chapter is based on the recently accepted work of the author:

D. A. Potoyan, P. I. Zhuravlev, G. A. Papoian; *J. Phys. Chem. B* (2012)

4.1 Introduction

It is well known that energy landscapes underlying protein functional dynamics may be rugged [197–200, 203, 210, 235, 236]. The protein dynamics on these landscapes involve the interconversion among myriad of conformational states that takes place on a spectrum of timescales from nanoseconds to seconds [203]. The roughness of the landscape is a result of many-body interaction between the protein chain, solvent molecules and ions. In many cases, the ruggedness means glassy dynamical behavior and lack of self-averaging, preventing a generalized statistical description of the functional dynamics [219, 237]. Therefore, computing detailed maps of the energy landscapes is indispensable for understanding mechanisms of functional processes of proteins, such as allosteric regulation [207, 237], native state dynamics [104, 197, 203, 236, 237], or biopolymer translocation [213]. One approach to mapping of the energy landscapes is to start from calculating the free energy

difference between two specific protein conformations, for example, two allosteric states. Then, in principle, one could extend this calculation to many more conformations of a protein, leading to a high resolution view into the energetic topography of the native basin.

Among a number of proteins used to gain insights into the interplay between protein dynamics and function, adenylate kinase (ADK) stands out as one of the most extensively studied, in the context of allosteric transitions. The structures of ADK's allosteric states, known as the open and closed forms, have been solved long ago [217]. Also the dynamics of conformational change in ADK has been a topic of numerous experimental [192,202,206,216,218,226,227,233] and theoretical [165,196, 205,211,215,232] studies. However, the free energy difference between the two main allosteric states of ADK has not yet been computed using rigorous approaches based on explicit solvent force fields. Meanwhile, this free energy difference is one of the important aspects underpinning the nature of ADK catalytic action. We calculate it in this work.

Extracting free energy differences from all atom simulations of proteins presents a formidable challenge. The difficulty largely stems from a large number of degrees of freedom which encumbers full sampling of the thermodynamic states. In a prior work, structurally based umbrella sampling free energy calculation for ADK was carried out with the implicit solvent [194]. As elaborated in a recent publication [238], the corresponding collective coordinate for umbrella sampling [194,195] used to compute conformational free energy differences has significant shortcomings, and may potentially lead to artifacts. A key problem of the widespread structural coordinates

like $\Delta RMSD$ or ΔQ, is the fact that these coordinates have unacceptably high degree of structural degeneracy. Namely, in these coordinates a single value maps into a substantial number of unrelated conformations, which defeats the whole purpose of measuring conformational free energy difference between two distinct states, since the very definition of states is inconsistent. To illustrate the structural degeneracy we have used a $\xi(Q_A, Q_B) = \Delta Q$ as a reaction coordinate to map conformational space of two structurally highly similar states of a small model protein (see Fig. 4.1). However, as one can see from the Fig 4.1, even in this favorably picked case the structural degeneracy is quite large, with states that are structurally dissimilar to both conformations lumped into the reference basins. To address this problem, we have recently developed a rigorous technique that permits calculation of conformational free energies between two arbitrary polymer conformations from explicit solvent all atom simulations, where by two conformations we mean two well defined regions of the polymer molecule phase space [238]. Our method largely consists of identifying a reaction coordinate which correlates with conformational transformation. Combining our coordinate with the enhance sampling simulations (such as umbrella sampling) provides a route for computing free energies or any other relevant equilibrium thermodynamic parameters regarding the transition. As demonstrated in the current work, this technique can be applied to large proteins, over two hundred residues, simulated in explicit solvent. Other approaches which attempt to accomplish the same goal use uncontrolled approximations [208, 230, 231] or create unphysical intermediate states [220] which makes their application to complex explicit solvent bimolecular systems rather challenging (see [238] for more detailed

discussion of alternative techniques).

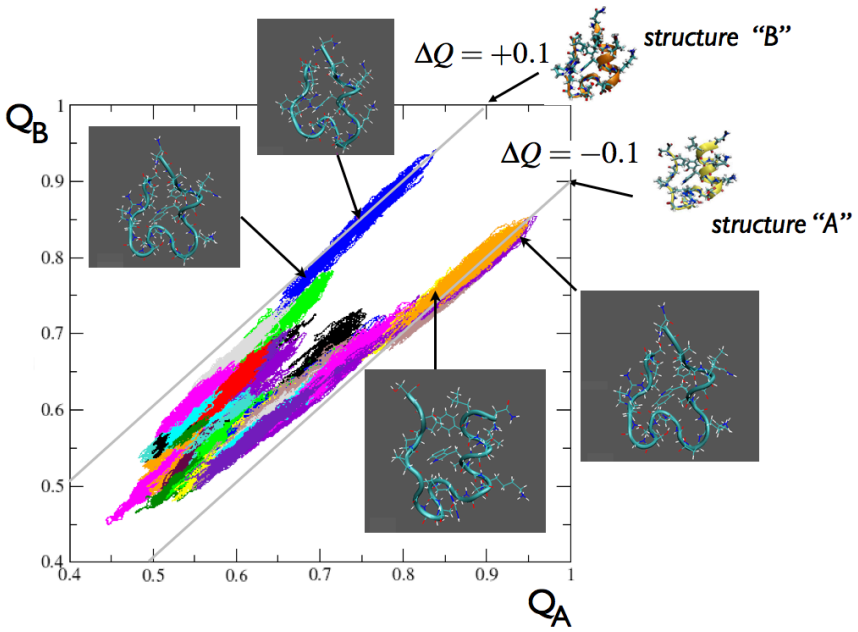

Figure 4.1: The illustration of a reaction coordinate degeneracy problem for defining conformational basins. On the plot are shown the conformational states of a model protein(trpcage, PDB ID 1L2Y) for which the $\Delta Q = Q_A - Q_B$ reaction coordinate was tested.

The proposed conformational reaction coordinate smoothly morphs one allosteric state into another, preserving structural locality near the allosteric basins [238]. The latter is a critical requirement missing in many commonly used alternative approaches. Locality means that if for a conformation X the value of the variable $\xi(X)$ is close to the value of the variable in the closed state ξ(closed), then X is structurally similar to the closed state. The same statement is also true for the open state [238]. Another important ingredient of the technique is the confinement potential which alleviates the problem of sampling the degrees of freedom transverse to the umbrella sampling variable. Its effect on the calculation amounts to enveloping

the dynamic trajectories of the system into a phase space "tube" without affecting the neighborhood of the two allosteric states. The tube prevents sampling of the phase space regions of high conformational entropy which contain unfolded and other non-relevant states, as it is not necessary in the case when only free energy difference between the two conformations is calculated. One of the goals of this chapter is to demonstrate that this new method, which was previously applied [238] to a 20-residue protein, Trp-Cage, straightforwardly scales up to an order-of-magnitude larger proteins.

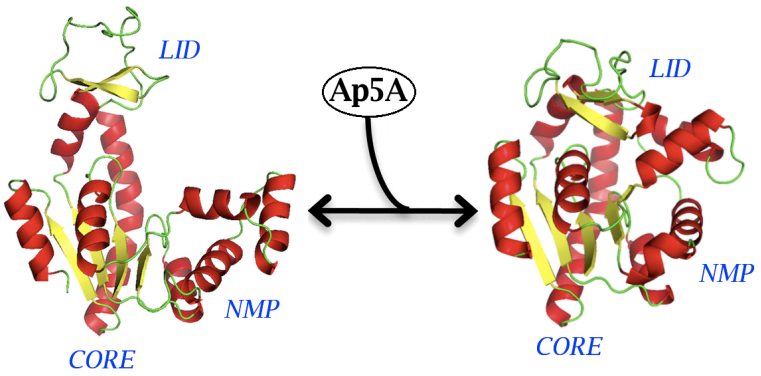

Figure 4.2: Crystallographic structures of open (4AKE) and closed (1AKE) forms of ADK are shown [217]

Adenylate kinase(ADK) is an important member of kinase family of proteins which catalyzes a key metabolic step of phosphoryl transfer between ADP molecules: $ADP \cdot Mg^{+2} + ADP \leftrightarrow ATP \cdot Mg^{+2} + AMP$. Structurally, ADK consists of three domains: LID, NMP and CORE [217]. The CORE domain comprises the bulk of the protein and is relatively static during allosteric transition, meanwhile NMP and LID form contacts with entering substrate by covering the bound substrate and preventing it from diffusing into the solvent environment. Many of the previous

computational studies focused on the mechanistic aspects of the collective domain motions, which is correlated with the enzymatic activity [226]. More specifically those studies identified the key rate limiting step of LID domain opening when bound to a ligand [193, 232].

In this work we find the free energy difference between the states in the absence of ligand. In significant contrast to the previously reported high (tens of k_BT) values of the free energy difference from implicit solvent simulations [194], we find that free energies of both states are comparable, with the difference on the order of ∼1–2 k_BT. In addition, we find that, quite unexpectedly, interfacial contacts between LID and NMP domains, characteristic of the closed state, start to form even when the conformation is close to the open state. In other words, the transition from the closed to the open basin might be characterized by a late transition state.

4.2 The free energy calculation technique

The central question addressed by our method [238] can be stated as follows: given two experimentally determined structures, A (open) and B (closed), how can we estimate the free energy difference ΔF_{AB} between them? Here we assume a time scale separation between fast sampling of the similar conformations around both open and closed states, and slow transitions between the states. Hence, the neighborhoods of the transition end points form corresponding basins of attraction (conformational basins), having a specific size depending on the molecular structure and inter-molecular interactions. Thus, the conformational free energy difference

that we are after is in fact the free energy difference between the conformational basins of A and B. The size of the basins should be chosen from the considerations external to our calculation technique.

The umbrella sampling variable (or reaction coordinate) $\xi(X)$ which is local near the allosteric states and continuously morphs one structure into another is described by the following functional form [238]:

$$\xi(X) = \exp\left[-\frac{(s(X,A) - s(A,B))^2 + (s(X,B) - 1)^2}{2\rho^2}\right] - \exp\left[-\frac{(s(X,B) - s(A,B))^2 + (s(X,A) - 1)^2}{2\rho^2}\right], \tag{4.1}$$

where $s(X,Y)$ is a measure of structural similarity between conformations X and Y, such as RMSD or fraction of shared contacts (Q). When overlap parameter Q is used in defining $s(X,Y)$, then $s(X,Y) = 0$ means X and Y have nothing in common, and $s(X,Y) = 1$ means it is the same conformation. The conformational phase space, therefore, maps into a square with sides graded from 0 to 1. Regarding the variable ξ (see Fig.4.3) defined in Eq. (4.1) as elevation above this square, it corresponds to a positive circular gaussian "peak" placed on the basin of conformation B and a negative circular gaussian "pit" placed on conformational basin A [238]. The ρ in the equation is a parameter which controls the resolution of ξ and has to be decided based on the specific aims of the problem (see Table.4.1) . If its value is too large, the resolution coarsens and might not capture the subtle structural differences in the respective basins leading to more or less random free energy estimates. On the

other hand making resolution too high at best increases the computational demand, by requiring more windows and higher values for spring constants to adequately sample the entire pathway. For $s(X, Y)$, we chose the structural overlap parameter widely employed in studies of protein folding and spin glasses [223], which is simply a generalized form of the fraction of shared contacts:

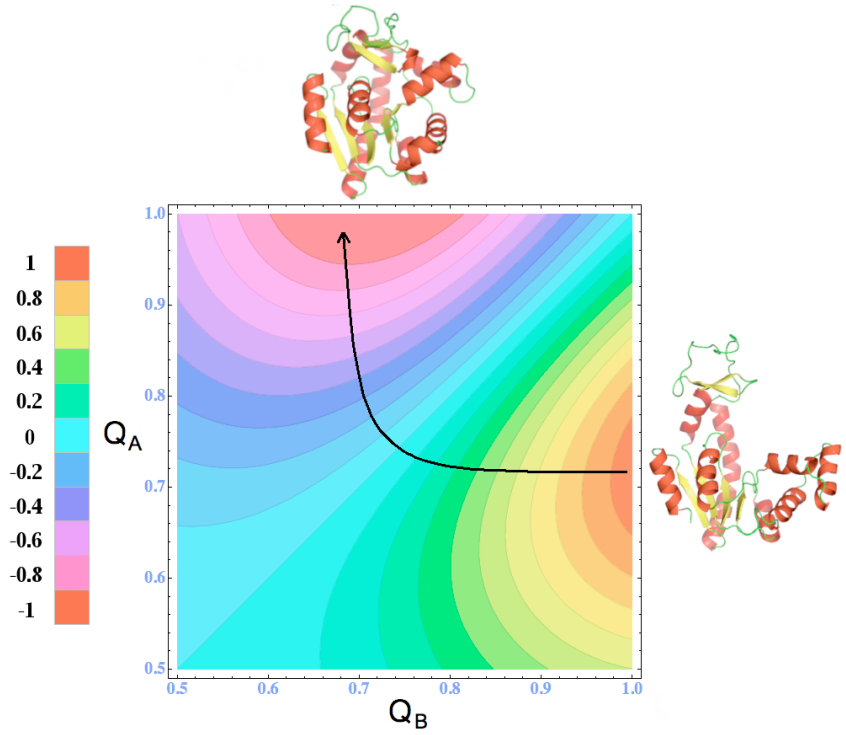

Figure 4.3: Contour plot of the reaction coordinate $\xi(X)$ [238], where X is an arbitrary point of the conformational space that maps into $Q_A(X)$ and $Q_B(X)$.The direction along which we sample conformations is indicated with a black arrow.

$$s(X, Y) = Q(X, Y) = \frac{1}{N} \sum_{i,j} \exp\left[-\frac{(r_{ij}^X - r_{ij}^Y)^2}{2\sigma^2} \right], \tag{4.2}$$

where σ sets the length-scale for the native contacts and is typically on the order of $C_\alpha - C_\alpha$ distance, e.g $\sim 1\mathring{A}$. Hereafter, for the notational simplicity we will avoid

the explicit indication of path variable X and instead will write the similarity as Q_A and Q_B.

The immense volume of the phase space orthogonal to umbrella sampling variable ξ with varying spectrum of relaxation timescales at different points also represents a problem. Whenever the relaxation of transverse fluctuations exceeds the simulation times the trajectory may fall in an "entropic trap" and potentially not reach equilibration and convergence.

To alleviate the difficulty of adequately sampling transverse fluctuations, we designed a confining potential V_c, which is added to the Hamiltonian and is meant to block the trajectory from escaping into high conformational entropy regions. Confinement potential V_c in $\mathcal{H}' = \mathcal{H} + V_c$ can be chosen such that it is equal to zero near the allosteric states, in which case the sought for free energy will not be affected by the virtue of the following equation

$$F'_B - F'_A = -\frac{1}{\beta} \ln \frac{\int\limits_{\Gamma_B} e^{-\beta \mathcal{H}'} \, \mathrm{d}\Gamma}{\int\limits_{\Gamma_A} e^{-\beta \mathcal{H}'} \, \mathrm{d}\Gamma} = -\frac{1}{\beta} \ln \frac{\int\limits_{\Gamma_B} e^{-\beta \mathcal{H}} \, \mathrm{d}\Gamma}{\int\limits_{\Gamma_A} e^{-\beta \mathcal{H}} \, \mathrm{d}\Gamma} \tag{4.3}$$

where $\beta = 1/k_B T$, Γ represents the full conformational space and the Γ_A and Γ_B are the portions of full space that correspond to conformational states of A and B (where $V_c = 0$) [238]. The particular form of confinement potential employed in the present work is chosen to have the following functional form (see Fig.4.4) which also illustrates the idea of phase space "tube":

$$V_c(Q_A, Q_B) = \varepsilon[2 - \tanh(\kappa((Q_A - a_1)(Q_B - a_1) - R_1)) + \tanh(\kappa((Q_A - a_2)(Q_B - a_2) - R_2))] \tag{4.4}$$

The potential on the (Q_A, Q_B) plane can be envisioned as two walls of height ε of hyperbolic shape (see Fig.4.4) with radii R_1 and R_2. The κ defines steepness of these walls (or hardness of the phase space tube enveloping the dynamic trajectory). a_1, a_2, are constants that determine positions of the hyperbolae. The optimal tube should allow many pathways connecting the open and closed states, but not excessively many. The particular form for the confining potential is not critical since it is eventually canceled when computing the free energy difference between conformations A and B. However, there are some requirements that limit the possible forms for the potential. For one thing, the confining potential should enclose the reference basins and ideally the actual transition pathway, but the latter is not necessary if one is after the free energy difference only. Secondly, the confining potential should have a smoothly rising "walls", which would push the trajectory within the confines of the tube without causing abrupt jumps in the trajectory. To find the shape of the tube we ran short (20–50 ps) simulations in each umbrella window, $U_i^{umb}(\xi) = k_{spring}(\xi(Q_A, Q_B) - \xi_i)^2$ without confinement, obtaining an approximation to a dominant transition pathway. The general shape is indicated by the regions sampled by the preliminary short trajectories which overall are localized in the upper right corner of the (Q_A, Q_B) plot (see Fig. 4.3). The hyperbolic shape given its simple algebraic form is a reasonable choice for the purpose of confining trajectories (irrespective of a protein) in the upper right corner by a smooth walls,

the height of which is controlled by the ε . One typically assigns some arbitrary high value for ε (e.g. 10.0 kcal/mol). The steepness κ on the other hand should be high enough (see Table.4.1) to prevent the trajectories from overcoming the wall, but low enough to not cause numerical instabilities when computing the forces near the boundaries.

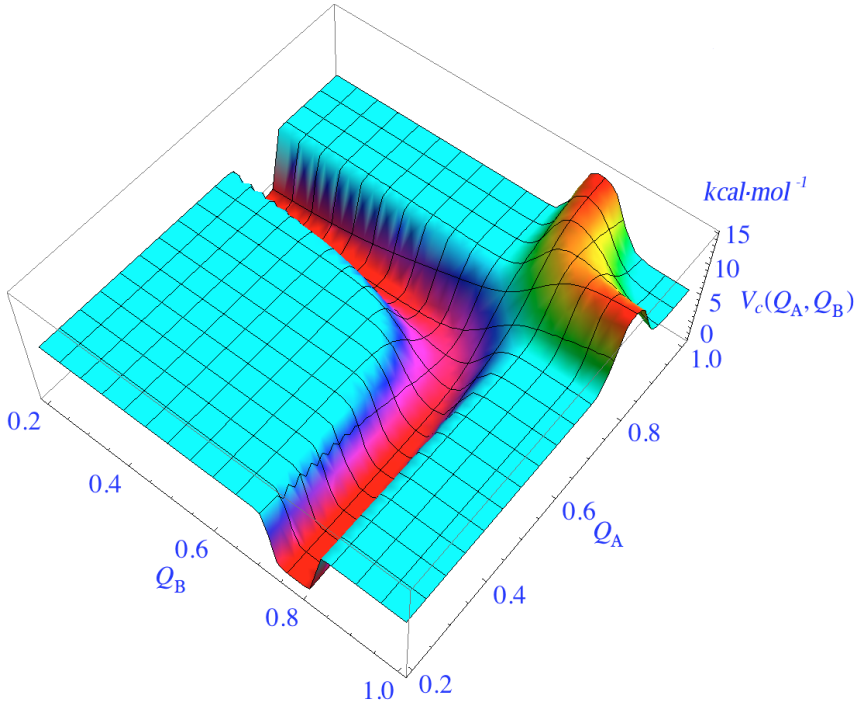

Figure 4.4: The potential V_c from Eq. (4.4) forms a "creek" confining the sampling trajectories inside. The shape of the "creek" is hyperbolic and the rise of its shores is hypertangential.

The tube parameters, a_1, a_2, R_1, R_2 (see Table.4.1), which determine the position and thickness of the tube, have to be chosen in such way as to leave the reference basins intact. Therefore, there is a minimally possible width of the tube. On the other hand, the width should not be too large, so the trajectories equilibrate on the timescales comparable to the length of the simulation and also allow for overlap

of trajectories in the neighboring windows. Thus, to determine the parameters we first run preliminary short simulations by gradually raising the thickness of the tube until the two of the mentioned conditions are satisfied.

4.3 Computational details

As starting structures we used atomic coordinates of the closed (1AKE) and open (4AKE) forms of ADK taken from the Protein Data Bank [217] (see Fig.4.2). After stripping off the ligand coordinates from the raw crystallographic structure of the closed state, we immersed both conformations in the TIP3P water boxes and added Na+ and Cl- ions to mimic the physiological concentration of a cell ($\sim$140 mM L^{-1}). Total number of ions and water was chosen the same for both systems. All subsequent simulations were performed with the program NAMD [221], utilizing CHARMM27 forcefield [212]. After initial minimization steps with the constrained and unconstrained protein coordinates, we heated the systems to 300K by performing 200 ps NVT simulations. NVT simulations were carried out in the contact to heat bath, emulated by Langevin dynamics with the friction coefficient of $4\,\mathrm{ps}^{-1}$. After heating steps we equilibrated the density of the system by performing 2 ns NPT simulation.

Fully equilibrated systems were then replicated among 121 windows. Approximately half of the umbrella sampling simulations were initiated from the closed form and the other half from the open form. We accumulated a series of $\sim$1.0 ns trajectories, carrying out two independent umbrella sampling simulations for each

Table 4.1: Parameter values used in the auxiliary potentials.

Parameter	Value
$Q(A,B)$	0.856
ρ	0.06
σ	1.5 $\mathring{A}$
k_{spring}	100 kcal/mol
ε	10 kcal/mol
κ	5000
a_1	0.76
a_2	0.88
R_1	0.0063
R_2	0.0044

window. The simulations with added umbrella potentials were performed with the LAMMPS software [222]. Before productive simulations, we ran multiple short trajectories in all windows in order to probe the local landscape and find optimal spring constants that guarantee sufficient overlap of reaction coordinate distributions. We have found that uniform spring constant values (see Table.4.1) across the equally spaced windows ($\delta\xi = 0.02$) sufficed to generate excellent overlap between umbrella sampling variable distributions of neighboring windows. The shape of the confining wall has been chosen such that it leaves the basins of the reference conformations unchanged meanwhile enveloping all the intermediate ones with a high walled tube.

Since there is a time scale separation of backbone motions and side-chain rotations we have included only C_α atoms in the definition of Q. We limited the summation in Q to only those pairs of C_α atoms that are closer than 12 Ångströms to each other in either of the conformations. Such definition of a contact has been used in the coarse grained protein folding studies [224]. By introducing distance cutoff in the definition of Q we significantly reduce the large number of conformational states which on average contribute equally to both basins and therefore act as a "noise".

4.4 Results and Discussion

To ascertain whether our simulations are capable of reproducing the sought free energy difference to a sufficient degree of accuracy we have performed two independent simulations using different sets of initial conditions, where the initial atomic velocities were randomized in all windows, and, hence, producing trajectories that are mutually unrelated. As one can see in Fig. 4.5, the resulting two free energy profiles are in semi-quantitative agreement, showing similar basic features. Another interesting feature of the obtained free energy profiles is the well pronounced minima near the endpoints of the reaction coordinate. These correspond to the basins of open and closed conformations of ADK. In case of ADK allosteric states, there are no external considerations which dictate the choice of the physically meaningful size of the conformational basins. Therefore, one could consider ADK conformations falling within the corresponding minima of the endpoints as belonging to one basin, leading to approximately $|\Delta\xi| = 0.2$ defining the basin size.

After defining the reference basins one can obtain the conformational free energy by integrating the potential of mean force along the reaction coordinate in respective basins (indicated using orange boxes in Fig. 4.5) according to

$$F_{\text{closed}} - F_{\text{open}} = -\frac{1}{\beta} \ln \frac{\int_{\Gamma_{\text{closed}}} e^{-\beta F(\xi)}\, \mathrm{d}\xi}{\int_{\Gamma_{\text{open}}} e^{-\beta F(\xi)}\, \mathrm{d}\xi}, \tag{4.5}$$

where $F(\xi)$ is the profile on Fig. 4.5. The free energy difference between the two allosteric states estimated from two uncorrelated trajectories was found to be $1.5 \pm 0.5\ k_B T$. This small value of free energy difference is in marked contrast with the work of Arora et al. [194] where the potential of mean force as a function of RMSD was estimated to be tens of $k_B T$ higher in free energy upon approaching the basin containing the ensemble of closed conformations.

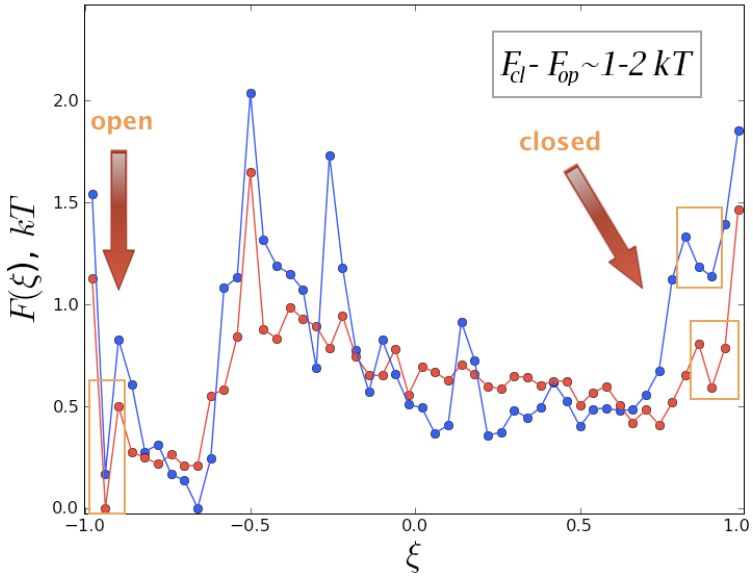

Figure 4.5: Free energy of ADK as a function of reaction coordinate ξ is plotted. Two different plots correspond to two completely independent simulations.

We suggest that the significant discrepancy between the current and previous

results may be explained by a combination of the following two factors, namely: 1) the implicit solvent that has been used in the earlier work is known for introducing a strong conformational bias [225], 2) the reaction coordinate used in the previous study is not local near the allosteric states, which may lead to artifacts [238]. The free energy profile computed in the current work is in semi-quantitative agreement with the single molecule FRET and NMR spin relaxation experiments which reported a relatively fast collective domain motions that take place on a nanosecond time-scale [226, 227] and a relatively rarer event of attaining catalytically competent closed conformation (e.g the one that is consistent with X ray structure for the closed state) on a microsecond time scale [233]. From the kinetics perspective based on Kramer's theory [234], if the closed state is reached within a microsecond or faster from the open state, then closed state's free energy (as well as the barrier for the transition) should be within several k_BT from the open state (assuming a pre-exponential factor which is characteristic of polypeptide chains [209]). NMR experiments with ligated ADK (achieved by excessive ATP concentration) also report the difference between the conformational basins to be on the order of 1 k_BT [218]. Finally, when coarse-grained computer simulations of ADK were parameterized by experimental data, the resulting difference in the free energies of open and closed states was also estimated to be 1-2 k_BT [211]

In addition, the obtained free energy profile offers a simple way of resolving the discrepancy concerning the dominant conformations that has been reported by NMR and single molecule FRET experiments. A few years ago, Hanson et al. [202] showed that in FRET experiments, closed-like conformations are sampled more

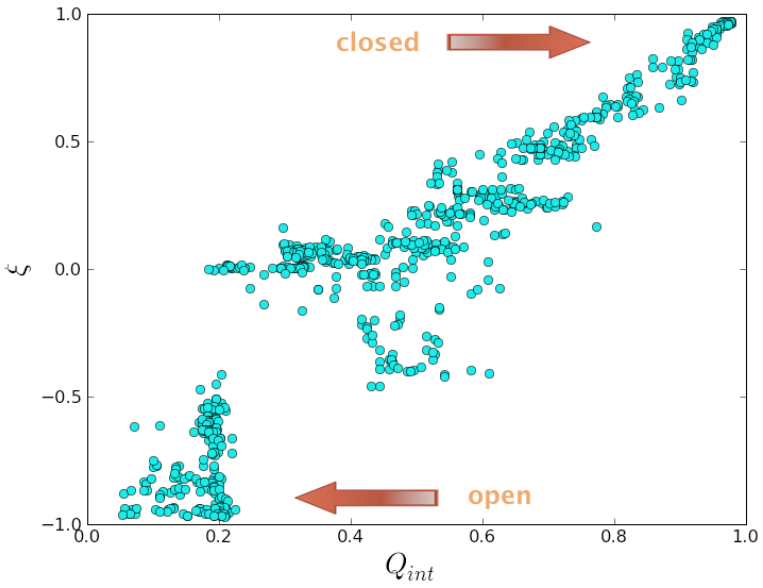

Figure 4.6: Each point on the plot corresponds to a snapshot from the simulation, for which a reaction coordinate ξ and a fraction of remaining interfacial contacts Q_{int} are calculated. Even close to the open state (in terms of ξ, and therefore structurally) many interfacial contacts are still present.

often. In contrast, the solution NMR suggests a picture [204, 227] where open-like conformations are heavily populated and transitions to the closed form occur occasionally. We explain the discrepancy in the following way: the conformational basin of the open state is narrower and well defined, while the basin for the closed state extends further along the reaction coordinate. Therefore, based on this line of reasoning, there is not a unique way to define the closed form of ADK, and other external criteria should be invoked to make the definition consistent with the problem. To clarify the latter statement, if we take a more generous definition of the closed basin, by extending the limit of integration in Eq.4.3 for the closed state

beyond the boundaries of the yellow box in Fig. 4.5, the free energy gap between closed and open states narrows considerably. We thus attribute the discrepancy between two different experimental techniques to the distance-sensitivity difference intrinsic to NMR and FRET that have skewed the conformational distribution in favor of one or the other form.

The transition path taken by the trajectories in our simulations may not correspond to the actual one (although the tube path was chosen based on short simulations which likely indicate some conformational preferences of the protein), but one can still infer useful information about the intermediate steps. For instance, the barrier in our profile provides the upper limit to the actual barrier, since cutting off parts of the phase space can only make some pathways unavailable to the trajectory, which, in turn, will remove the contribution of these cut-off pathways to lowering the free energy of the transition state.

Apart from the thermodynamics, one can gain further insight into the nature of conformational transition by analyzing the structural evolution across the umbrella sampling windows. Comparing pairwise interatomic distances in reference states, we have identified the evolution of *interfacial contacts*, the contacts that are being broken when going from closed to the open form on the interface of LID and NMP domains. Using the interfacial contacts we have computed Q_{int} for all conformations in each window and plotted the distribution of Q_{int} values against the order parameter, ξ, that has been used in the free energy simulations (see Fig. 4.6). From the plot one can see the nonlinear dependence of the order parameter on the fraction of native contacts, with an abrupt transition that coincides with the location of the

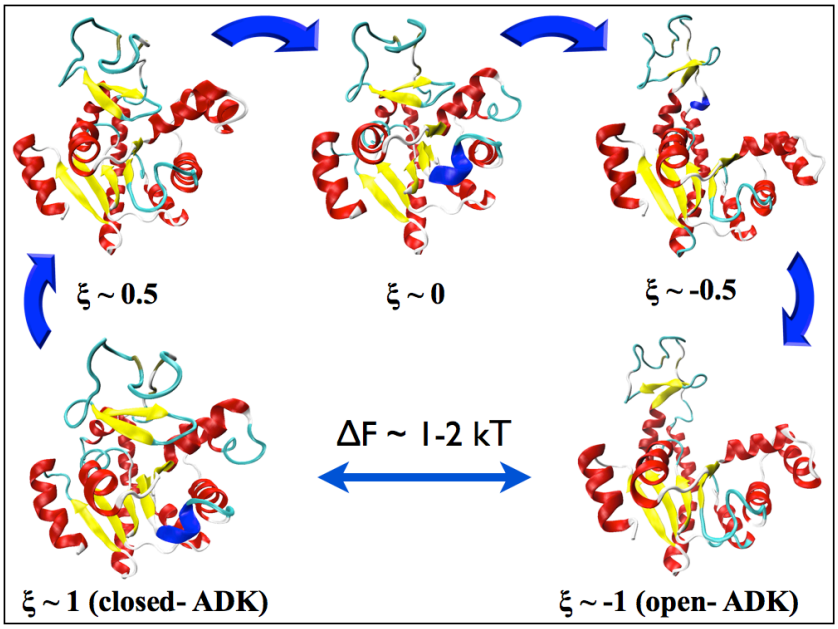

Figure 4.7: The intermediate structures along the pathway, selected from the appropriate windows. The transition from the closed to the open state goes through many closed like states followed by an abrupt opening of the NMP domain in the later stages.

barrier on Fig.4.5.

On the other hand, a naïve expectation for the open to close transition would be that of an abrupt disruption of contacts when the LID-NMP domains smoothly and linearly separate from each other by a few Ångströms. However, contrary to what one would expect, a significant fraction of the interfacial contacts are retained even when the system has moved further away from the basin of the closed state (see Figs. 4.6 and 4.7). This observation might be an artifact from confinement potential, but about 20% of interfacial contacts are retained close to the open state, where the influence of the confinement potential is low. This relation between the

structure and interfacial contacts means that conformational change is structurally more intricate, involving partial breakage of contacts taking place gradually or in several stages.

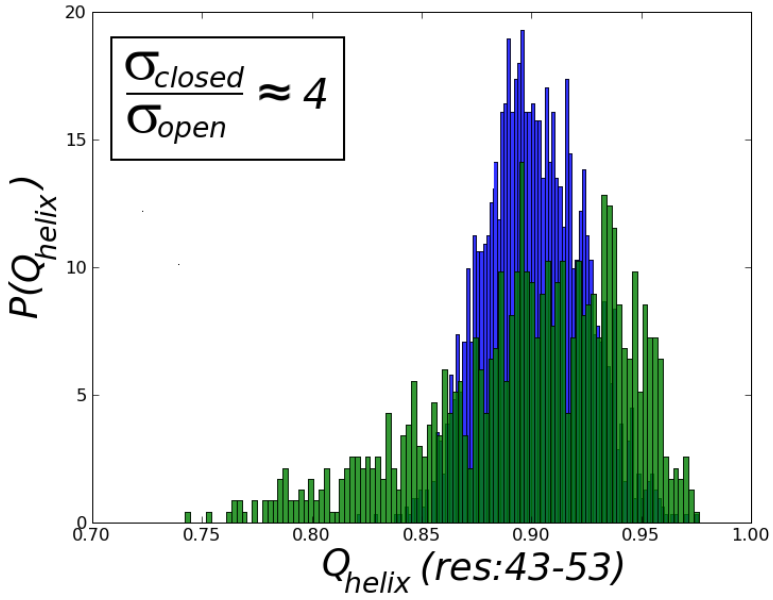

Figure 4.8: The evidence for the local frustration in the closed form of the ADK. The blue and green histograms are the distribution of Q_{helix} for the closed and the open forms respectively. The $\sigma(open)$ and $\sigma(closed)$ denote the standart deviation in the respective distributions.

The more shallow landscape near the closed state ($\xi = 0 - 1.0$) translates into a higher conformational entropy for the ensemble of the closed-like conformations, which are qualitatively defined as the collection of states that have higher structural similarity to the reference closed state. This somewhat unexpected result is supported by the results of our local "structural variability" calculations in the α-helices and the B factor distributions [216] measured for the crystal structures of the two

forms of the ADK. We quantify the so called local "structural variability" by the distribution of the Q values (see Eq. 4.2) in the definition of which we include only the residues that are part of the alpha helix. We denote the structural variability of a helix as Q_{helix}. The broader distributions of Q_{helix} implies higher "structural variability" and hence higher conformational entropy for the particular α helical segment and vice versa. By computing the distribution of Q_{helix} in the basins of an open ($\xi \approx -1$) and the closed states ($\xi \approx 1$) we noticed that overall the helices in the closed from are characterized by a lower structural variability compared to the open form (see Fig. 4.8). In particular, the alpha helix composed of residues: 43-53 in the NMP domain showed striking contrast in terms of the "structural variability" of the two forms. Approximately half of the indicated residues are in a disordered state in the closed form of the ADK (according to the DSSP measure) and hence show a much higher structural variability compared to the ordered form, (see Fig 4.8) where the same residues are all part of an α helix. From the protein physics viewpoint this observation can be explained by the local frustration in the closed form of the ADK, where the multiple contacts between the domains are made at the expense of some of the contacts which stabilize the helices in the respective domains. From the thermodynamic point of view, one may entertain the idea that after the domain closure the enthalpy gained from forming the contacts is is not completely dissipated into the solvent environment but is partially channeled into the helices by increasing their structural variability. This essentially amounts to an "entropic transfer" in the ADK from the inter-domain flexibility of the open form to the intra-helix disorder in the closed form. The net effect for the entropic transfer

is lowering the free energy gap, which is beneficial for the allostery and the enzyme recycling.

Other native contacts might be more significant to the transition than the interfacial ones. The retention of the interfacial contacts near the open state prompts one to speculate that the transition state ensemble is more likely to be located closer to the open state. From the kinetic point of view this would be an indication of an anti-Hammond type behavior: more structural reorganization is needed to reach the transition state, if one starts from the closed, than from the open state, while the free energy increase is, on the contrary, larger if one starts from the open state [201,214].

4.5 Concluding remarks

The free energy difference between the open and closed forms of ADK in the absence of ligands, computed in this work, turned out to be rather small ($\sim$1–2 k_BT). Two fully independent calculations indicate that the results are reasonably converged and reproducible. From the biological perspective, the relatively small free energy difference between the allosteric states may facilitate fine control of allosteric transition by environmental perturbations and signaling. In addition, we found that even when some of the NMP-LID interfacial contacts are formed, the typical conformations are still structurally more similar to the open form, suggesting a late transition state for domain opening. Our structural variability calculations further clarify the mechanistic and thermodynamic signature of the transition. We have found that in closed state domains have higher structural variability compared

to the open form which acts as a counterweight balancing the entropic loss associated with domain closure. We rationalize these observation by providing a mechanism of entropic transfer, which is a way for the allosteric transition to lower its free energy gap between the endpoint states.

We conclude that the recently developed method for calculating conformational free energy differences [238] can be applied to systems of real biological importance, for instance, large proteins like adenylate kinase. In order to calculate a free energy profile corresponding to the actual kinetics of the allosteric transition, the confinement tube trajectory should be dynamically updated, instead of being statically defined, allowing the system to find the dominant path of minimum free energy. The latter goal will be pursued in our subsequent studies.

Appendix A

Appendix for Chapter 2

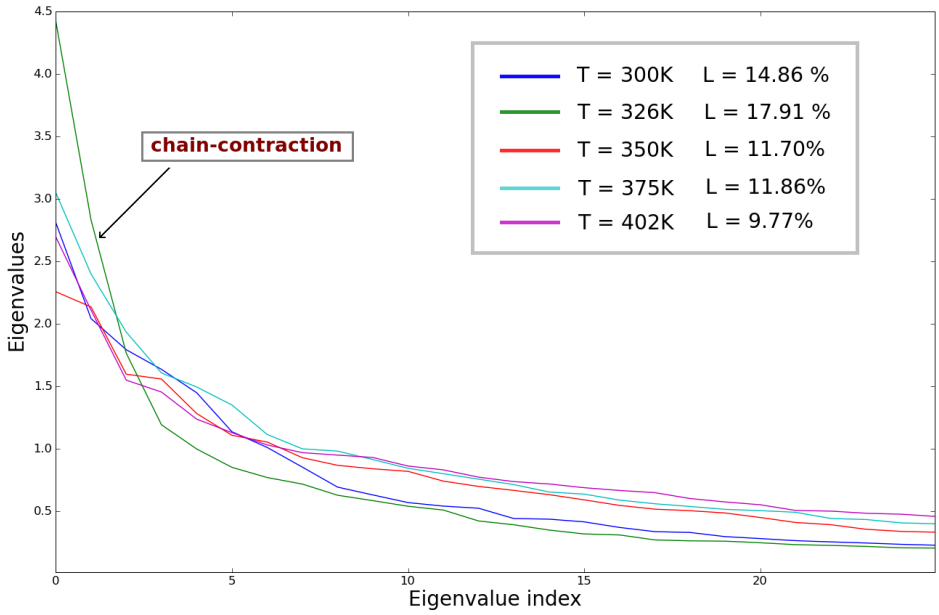

Figure A.1: Eigenvalues obtained from dPCA of the H4 tail dynamics is plotted against the corresponding eigenvalue indices, indicating faster decay at moderately elavated temperature of 326 K, due to chain contraction. The relative contributions of the first two PCs to overall dynamics at different temperatures are indicated with letter "L" in the legend box.

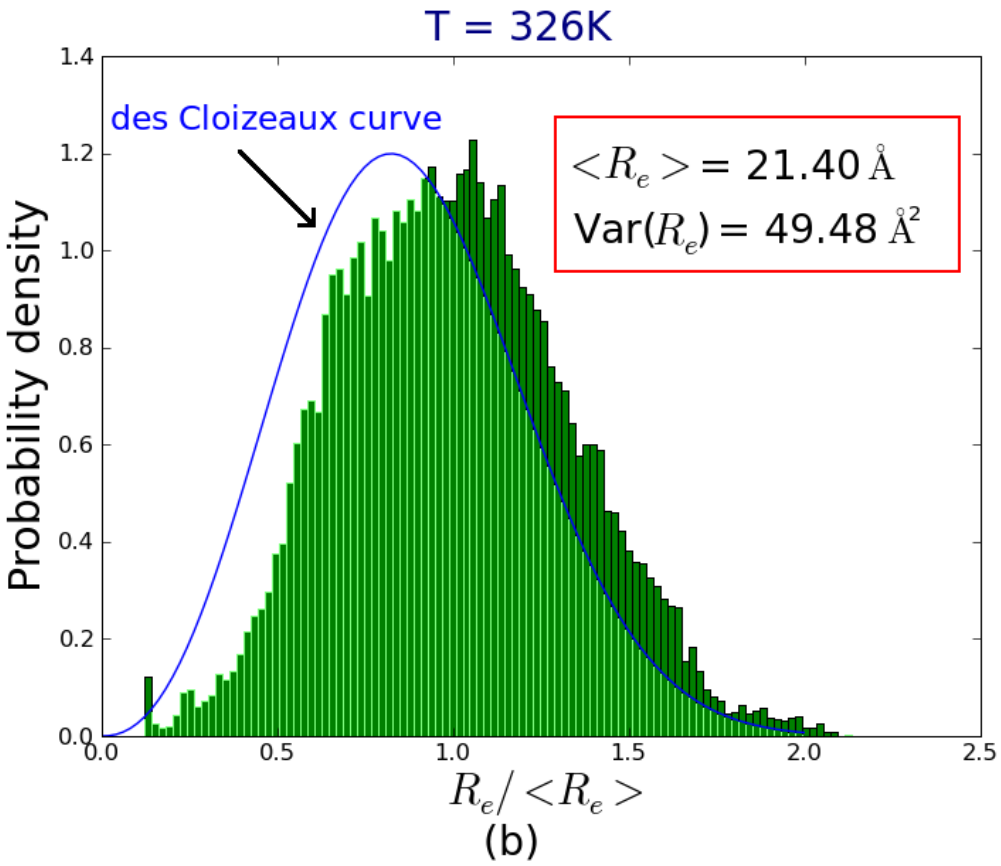

Figure A.2: End-to-end distance histogram of the H2A tail is shown at T=326 K.

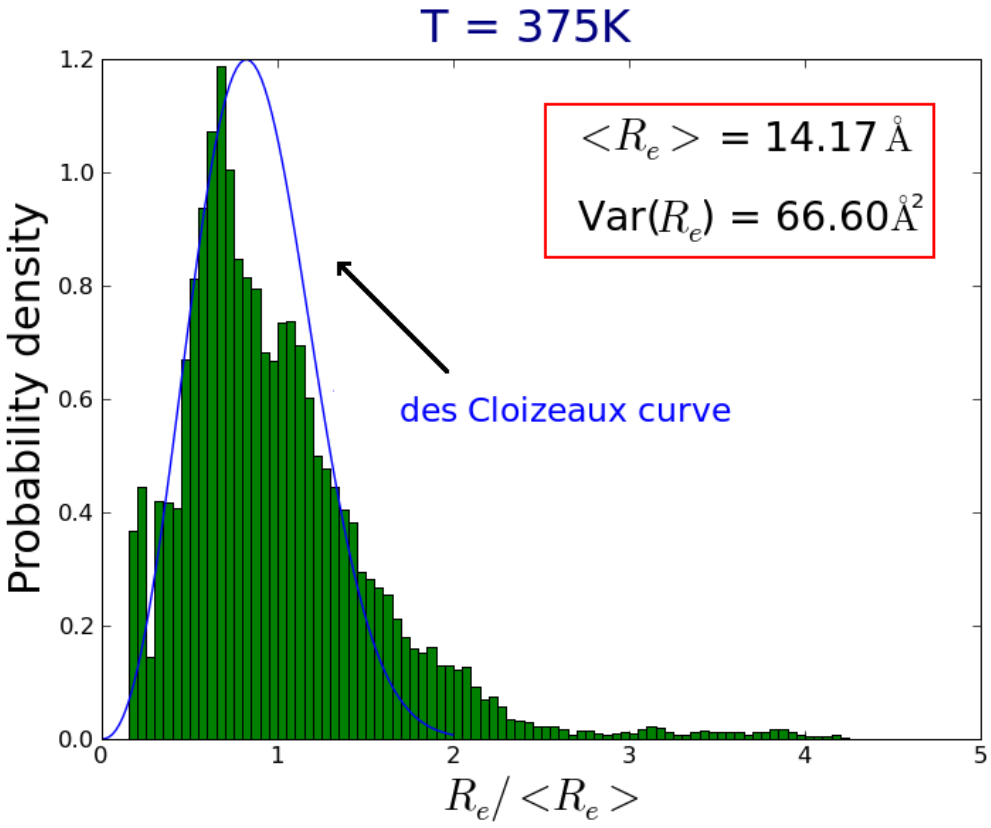

Figure A.3: End-to-end distance histogram of the H4 tail is shown at T=375 K.

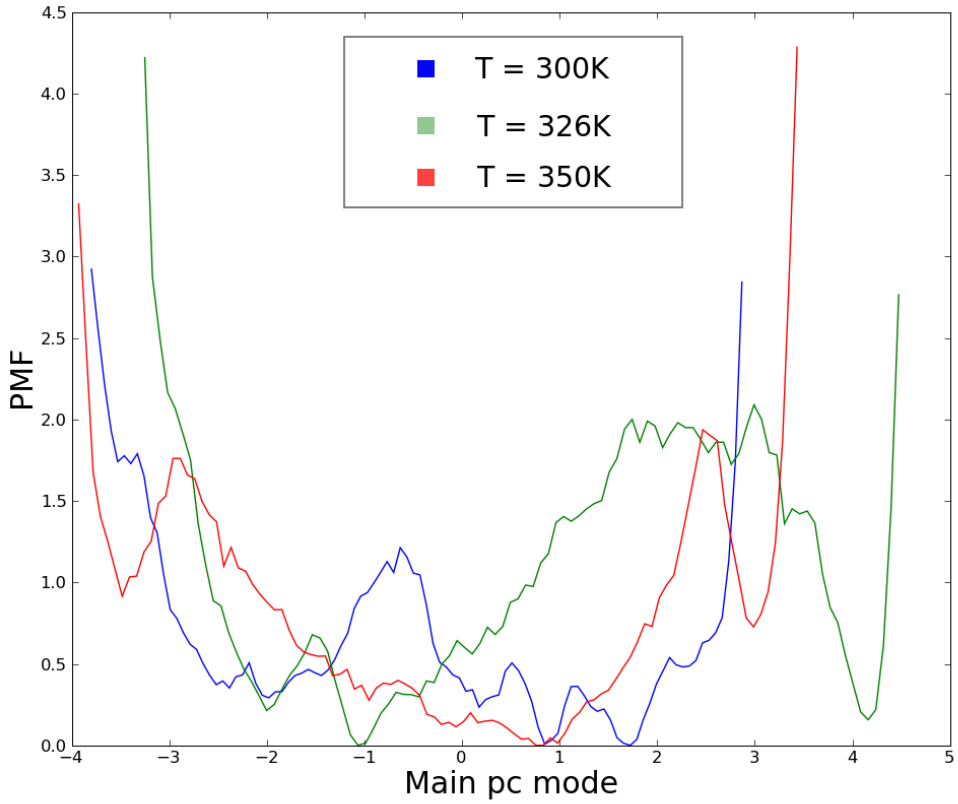

Figure A.4: Free energy projection of the H4 tail dynamics at different temperatures into the main PC mode, showing the appearance of a new state at T=326 K, corresponding to the conformations of the contracted chain and its subsequent destabilization at T= 350 K.

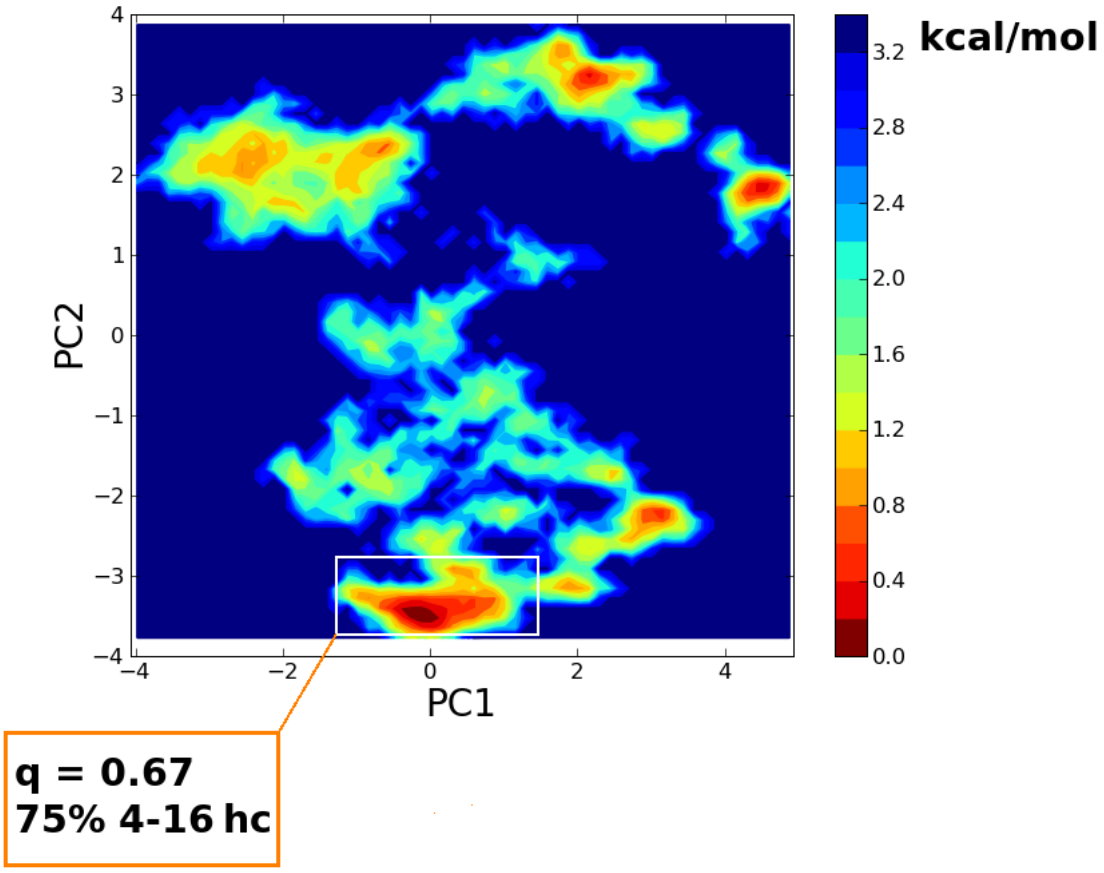

Figure A.5: Free energy projection of the H3 tail dynamics at 300 K into its two main principal components is shown, based on Eq. 3.3 of the main text.

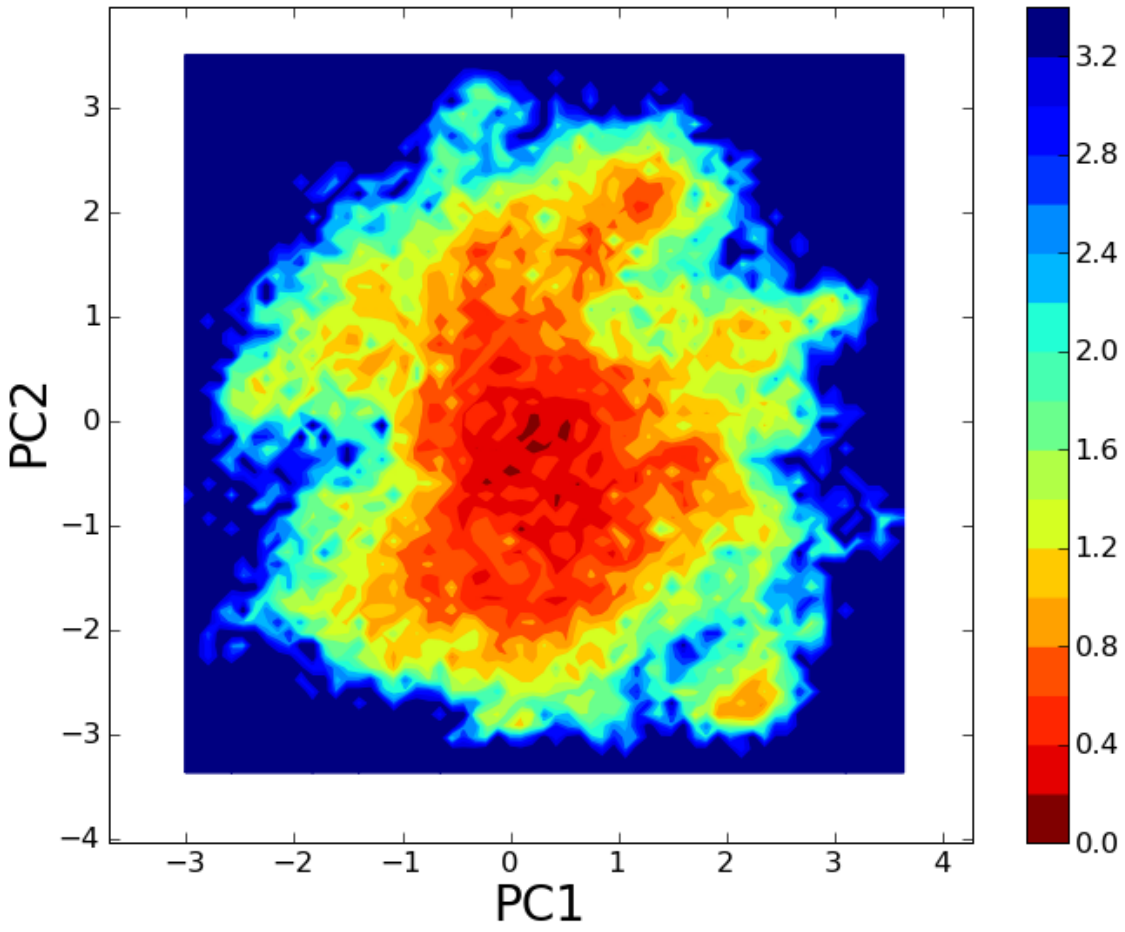

Figure A.6: Free energy projection of the H4 tail dynamics at 375 K into its two main principal components is shown, based on Eq. 3.3 of the main text.

Bibliography

[1] J Venter and et al. The sequence of the human genome. *Science*, 291 1304, 2001.

[2] N Gilbert and et al. Chromatin architecture of the human genome Gene-rich domains are enriched in open chromatin fibers. *Cell*, 118 555, Jan 2004.

[3] T Misteli. Beyond the sequence cellular organization of genome function. *Cell*, 128 787, 2007.

[4] S Baylin and Kornel E Schuebel. Genomic biology the epigenomic era opens. *Nature*, 448 548, Jan 2007.

[5] Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. *Molecular Biology of the Cell*. Garland Science, 5 edition, 11 2007.

[6] P. Pincus H. Schiessel. Counterion-condensation-induced collapse of highly charged polyelectrolites. *Macromolecules*, 31 7953-7959, 1998.

[7] Davey CA, Sargent DF, Luger K, Maeder AW, Richmond TJ. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. *J. Mol. Biol.*, 319 1097, 2002

[8] Peter J. Horn and Craig L. Peterson Chromatin Higher Order Folding: Wrapping up Transcription. *J. Biol. Chem.*, 297 33701, 2002

[9] F Gordon, K Luger and Jeffrey C. Hansen. The Core Histone N-terminal Tail Domains Function Independently and Additively during Salt-dependent Oligomerization of Nucleosomal Arrays. *J. Biol. Chem.*, 280 33701, 2005.

[10] K. Maeshima, S. Hihara; M. Eltsov Chromatin structure does the 30-nm fibre exist in vivo? *Curr. Op. Cell. biol.*, 22 1-7, 2010.

[11] A J Andrews and Karolin Luger. Nucleosome structure(s) and stability variations on a theme. *Annu Rev Biophys*, 40 99-117, Jun 2011.

[12] Eric I Campos and Danny Reinberg. Histones annotating chromatin. *Annu Rev Genet*, 43 559-99, Jan 2009.

[13] Anton Eberharter and Peter B Becker. Histone acetylation a switch between repressive and permissive chromatin. second in review series on chromatin dynamics. *EMBO Rep*, 3(3) 224-9, Mar 2002.

[14] Guohong Li and Danny Reinberg. Chromatin higher-order structures and gene regulation. *Curr Opin Genet Dev*, 21(2) 175-86, Apr 2011.

[15] Tamaki Suganuma and Jerry L Workman. Signals and combinatorial functions of histone modifications. *Annual review of biochemistry*, 80 473-99, Jun 2011.

[16] V Bloomfield. Dna condensation. *Curr. Opin Struc. Biol.*, 6 334, 1996.

[17] N. D. Socci, J. N. Onuchic, P. G. Wolynes Diffusive dynamics of the reaction coordinate for protein folding funnels *J. Chem. Phys.*, 104 5860, 1996

[18] Samuel S. Cho, Yaakov Levy, Peter G. Wolynes P versus Q: Structural reaction coordinates capture protein folding on smooth landscapes *Proc. Natl. Acad. Sci. USA*, 103 586, 2006

[19] R. Zhou Trp-cage: Folding free energy landscape in explicit water *Proc. Natl. Acad. Sci. USA*, 100 13280, 2003

[20] Dekker J, Rippe K, Dekker M, Kleckner N Capturing chromosome conformation *Science* , 1306, 295 2002

[21] E. Lieberman-Aiden et al. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome *Science* , 326, 289 2009

[22] Dostie J, Dekker J Mapping networks of physical interactions between genomic elements using 5C technology *Nat. Protoc.* , 2, 988 2007

[23] E. Kim, S. Jang, Manho Lim, Y. Pak Free Energy Landscape of the FBP28 WW Domain by All-Atom Direct Folding Simulation *J. Phys. Chem. B*, 114, 7686, 2010

[24] A. Pohorille, C. Jarzynski, C. Chipot Good Practices in Free-Energy Calculations *J. Phys. Chem. B*, 114, 10235 2010

[25] J. M. Rickman, R. LeSar Free-Energy Calculations in Materials Research *Annu. Rev. Mater. Res.* , 32, 195 2002

[26] Y Fukunishi, D. Mitomo, H. Nakamura Protein-Ligand Binding Free Energy Calculation by the Smooth Reaction Path Generation (SRPG) Method *J. Phys. Chem. B*, 49, 1944 2009

[27] H-J. Woo, B. Roux Calculation of absolute proteinligand binding free energy from computer simulations *Proc. Natl. Acad. Sci. USA*, 102 6825, 2005

[28] A. M. Ferrenberg and R. H. Swendsen New Monte Carlo technique for studying phase transitions *Phys. Rev. Lett.*, 61 2635, 1988

[29] M Cecchini and M Karplus. Conformational free-energy difference of a miniprotein from nonequilibrium simulations. *J. Phys. Chem. Lett.*, 1, 1922 2010.

[30] V. Spiwok, P. Lipovová, B. Králová. Metadynamics in essential coordinates: Free energy simulation of conformational changes. *J. Phys. Chem. B*, 111(12) 3073-3076, 2007.

[31] L. Zheng M. Chen, W. Yang Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems. *Proc. Natl. Acad. Sci.*, 105, 2022720232 2008.

[32] J.D Chodera et al. Use of the Weighted Histogram Analysis Method for the Analysis of Simulated and Parallel Tempering Simulations *J. Chem. Theory Comput.*, 3, 26 2007.

[33] Janke, W. Statistical analysis of simulations: Data correla- tions and error estimation in *Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms*; Grotendorst, J., Marx, D., Murmatsu, A., Eds.; John von Neumann Institute for Computing vol 10 2002.

[34] G.M. Torrie, J.P. Valleau Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling *J. Comp. Phys*, 23, 187 1977.

[35] J. Valleau and D. Card Monte Carlo Estimation of the Free Energy by Multistage Sampling *J. Chem. Phys.*, 57, 5457 1972.

[36] J. Kastner Umbrella sampling *WIREs Comp. Mol. Sci.* , 1, 932 2011

[37] E. A. Carter, G. Ciccotti, J. T. Hynes, and R. Kapral Umbrella sampling *Chem. Phys. Lett.* , 156, 472 1989

[38] Zwanzig, R. W. Umbrella sampling *J. Chem. Phys.* , 22, 1420 1954

[39] Mark E. Tuckerman *Statistical mechanics: theory and molecular simulation* . Oxford University Press 2010

[40] J. Kstner, W. Thiel Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: 'Umbrella Integration *J. Chem. Phys.* , 123, 144104 2005

[41] Shirts MR and Chodera JD Statistically optimal analysis of samples from multiple equilibrium states *J. Chem. Phys.* , 129, 124105 2008

[42] Temperature generator for replica exchange molecular dynamics simulations *http //folding.bmc.uu.se/remd/*, 2007

[43] *D.A. Case, et al., (2008), AMBER 10, University of California, San Francisco.*, 2008.

[44] et al. A. K. Dunker. Intrinsically disordered protein. *J Mol Graph Model*, 19(1) 26-59, 2001.

[45] M Rubinstein A. V. Dobrynin, R.H. Colby. Scaling theory of polyelectroliyte solutions. *Macromolecules*, 28 1859-1871, 1995.

[46] E. Schrodinger *What is Life?*. Cambridge University Press 1992

[47] Alexandros Altis, Phuong H Nguyen, Rainer Hegger, and Gerhard Stock. Dihedral angle principal component analysis of molecular dynamics simulations. *J Chem Phys*, 126(24) 244111, 2007.

[48] Alexandros Altis, Moritz Otten, Phuong H Nguyen, Rainer Hegger, and Gerhard Stock. Construction of the free energy landscape of biomolecules via dihedral angle principal component analysis. *J Chem Phys*, 128(24) 245102, 2008.

[49] L. Marino-Ramirez. The Histone Database: an integrated resource for histones and histone fold-containing proteins *Database, Vol. 2011*, 2011.

[50] G. Arya and T. Schlick. Role of histone tails in chromatin folding revealed by a mesoscopic oligonucleosome model. *Proc Natl Acad Sci USA*, 103(44) 16236-41, 2006.

[51] G. Arya, Q. Zhang, and T. Schlick. Flexible histone tails in a new mesoscopic oligonucleosome model. *Biophysical Journal*, 91(1) 133-50, 2006.

[52] Gaurav Arya and Tamar Schlick. Role of histone tails in chromatin folding revealed by a mesoscopic oligonucleosome model. *Proc Natl Acad Sci U S A*, 103(44) 16236-16241, 2006.

[53] Gaurav Arya and Tamar Schlick. A tale of tails how histone tails mediate chromatin compaction in different salt and linker histone environments. *J Phys Chem A*, 113(16) 4045-4059, 2009.

[54] Juan Ausi and D. Wade Abbott. The many tales of a tail carboxyl-terminal tail heterogeneity specializes histone h2a variants for defined chromatin function. *Biochemistry*, 41(19) 5945-5949, 2002.

[55] J. L. Banres, A. Martin, and J. Parello. The n tails of histones h3 and h4 adopt a highly structured conformation in the nucleosome. *J Mol Biol*, 273(3) 503-508, 1997.

[56] Robert B Best, Nicolae-Viorel Buchete, and Gerhard Hummer. Are current molecular dynamics force fields too helical? *Biophys J*, 95(1) L07-L09, 2008.

[57] Kerstin Bystricky, Patrick Heun, Lutz Gehlen, Jrg Langowski, and Susan M Gasser. Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques. *Proc Natl Acad Sci U S A*, 101(47) 16495-16500, 2004.

[58] David Chandler. *Introduction to Modern Statistical Mechanics*. OXFORD UNIVERSITY PRESS, Oxford, New York, 1987.

[59] A. Chepelianskii, F. Mohammad-Rafiee, E. Trizac, and E. Raphal. On the effective charge of hydrophobic polyelectrolytes. *J Phys Chem B*, 113(12) 3743-3749, 2009.

[60] Shirish M Chitanvis. Theory of polyelectrolytes in solvents. *Phys Rev E Stat Nonlin Soft Matter Phys*, 68(6 Pt 1) 061802, 2003.

[61] Scott L Crick, Murali Jayaraman, Carl Frieden, Ronald Wetzel, and Rohit V Pappu. Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions. *Proc Natl Acad Sci U S A*, 103(45) 16764-16769, 2006.

[62] F. Cubizolles and S. M. Gasser. The nucleosome from wallflower to queen of the ball. *Genome Biol*, 2(10) R4023, 2001.

[63] Weiwei Dang, Kristan K Steffen, Rocco Perry, Jean A Dorsey, F. Brad Johnson, Ali Shilatifard, Matt Kaeberlein, Brian K Kennedy, and Shelley L Berger.

Histone h4 lysine 16 acetylation regulates cellular lifespan. *Nature*, 459(7248) 802-807, 2009.

[64] J. des Cloizeaux. Lagrangian theory for a self-avoiding random chain. *Phys. Rev. A*, 10(5) 1665-1669, 1974.

[65] Feng Ding, Ramesh K Jha, and Nikolay V Dokholyan. Scaling behavior and structure of denatured proteins. *Structure*, 13(7) 1047-1054, 2005.

[66] Michael F Dion, Steven J Altschuler, Lani F Wu, and Oliver J Rando. Genomic characterization reveals a simple histone h4 acetylation code. *Proc Natl Acad Sci U S A*, 102(15) 5501-5506, 2005.

[67] Benedetta Dorigo, Thomas Schalch, Kerstin Bystricky, and Timothy J Richmond. Chromatin fiber folding requirement for the histone h4 n-terminal tail. *J Mol Biol*, 327(1) 85-96, 2003.

[68] Benedetta Dorigo, Thomas Schalch, Alexandra Kulangara, Sylwia Duda, Rasmus R Schroeder, and Timothy J Richmond. Nucleosome arrays reveal the two-start organization of the chromatin fiber. *Science*, 306(5701) 1571-1573, 2004.

[69] A. K. Dunker, J. D. Lawson, C. J. Brown, R. M. Williams, P. Romero, J. S. Oh, C. J. Oldfield, A. M. Campen, C. M. Ratliff, K. W. Hipps, J. Ausio, M. S. Nissen, R. Reeves, C. Kang, C. R. Kissinger, R. W. Bailey, M. D. Griswold, W. Chiu, E. C. Garner, and Z. Obradovic. Intrinsically disordered protein. *J Mol Graph Model*, 19(1) 26-59, 2001.

[70] A. Keith Dunker, Israel Silman, Vladimir N Uversky, and Joel L Sussman. Function and structure of inherently disordered proteins. *Curr Opin Struct Biol*, 18(6) 756-764, 2008.

[71] Robert N Dutnall. Cracking the histone code one, two, three methyls, you're out! *Mol Cell*, 12(1) 3-4, 2003.

[72] Nicolas L Fawzi, Aaron H Phillips, Jory Z Ruscio, Michaeleen Doucleff, David E Wemmer, and Teresa Head-Gordon. Structure and dynamics of the abeta(21-30) peptide from the interplay of nmr experiments and molecular simulations. *J Am Chem Soc*, 130(19) 6145-6158, 2008.

[73] Maxim V Fedorov, Jonathan M Goodman, and Stephan Schumm. To switch or not to switch The effects of potassium and sodium ions on alpha-poly-l-glutamate conformations in aqueous solutions. *J Am Chem Soc*, 131(31) 10854-10856, 2009.

[74] Mark S Formaneck, Liang Ma, and Qiang Cui. Effects of temperature and salt concentration on the structural stability of human lymphotactin insights from molecular simulations. *J Am Chem Soc*, 128(29) 9506-9517, 2006.

[75] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes. The energy landscapes and motions of proteins. *Science*, 254(5038) 1598-1603, 1991.

[76] Debabani Ganguly and Jianhan Chen. Atomistic details of the disordered states of kid and pkid. implications in coupled binding and folding. *J Am Chem Soc*, 131 5214-5223, 2009.

[77] Garca. Large-amplitude nonlinear motions in proteins. *Phys Rev Lett*, 68(17) 2696-2699, 1992.

[78] Ha and Thirumalai. Conformations of a polyelectrolyte chain. *Phys Rev A*, 46(6) R3012-R3015, 1992.

[79] J. C. Hansen, C. Tse, and A. P. Wolffe. Structure and function of the core histone n-termini more than meets the eye. *Biochemistry*, 37(51) 17637-17641, 1998.

[80] Jeffrey C Hansen. Conformational dynamics of the chromatin fiber in solution determinants, mechanisms, and functions. *Annu Rev Biophys Biomol Struct*, 31 361-392, 2002.

[81] Viktor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins*, 65 712, 2006.

[82] Pai-Yi Hsiao and Erik Luijten. Salt-induced collapse and reexpansion of highly charged flexible polyelectrolytes. *Phys Rev Lett*, 97(14) 148301, 2006.

[83] Veerasamy Jayaraj, Ramamoorthi Suhanya, Marimuthu Vijayasarathy, Perumal Anandagopu, and Ekambaram Rajasekaran. Role of large hydrophobic residues in proteins. *Bioinformation*, 3(9) 409-412, 2009.

[84] Giovanni Ciccotti Jean-Paul Ryckaert and Herman J. C. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23 327-341, 1977.

[85] Pu-Yeh Kan, Tamara L Caterino, and Jeffrey J Hayes. The h4 tail domain participates in intra- and internucleosome interactions with protein and dna

during folding and oligomerization of nucleosome arrays. *Mol Cell Biol*, 29(2) 538-546, Jan 2009.

[86] Hidenori Kato, James Gruschus, Rodolfo Ghirlando, Nico Tjandra, and Yawen Bai. Characterization of the n-terminal tail domain of histone h3 in condensed nucleosome arrays by hydrogen exchange and nmr. *J Am Chem Soc*, 131 15104-15105, 2009.

[87] Koji Nemoto Koji Hukushima. Exchange monte carlo method and application to spin glass simulations. *Journal of the Physical Society of Japan*, 65(6) 1604-1608, 1996.

[88] Nikolay Korolev, Alexander P Lyubartsev, and Lars Nordenskild. Computer modeling demonstrates that electrostatic attraction of nucleosomal dna is mediated by histone tails. *Biophys J*, 90(12) 4305-4316, 2006.

[89] Shankar Kumar, Djamal Bouzida, Robert H. Swendsen, Peter A. Kollman, and John M. Rosenberg. The weighted histogram analysis method for free-energy calculations on biomolecules. i The method. *J. Comput. Chem.*, 13(8) 1011-1021, 1992.

[90] A. M. Swendsen, R. H. Swendsen, Optimized Monte Carlo Data Analysis *Phys. Rev. Lett*, 63 1195, 1989.

[91] Edited by, C. Chipote, A Pohorille Free Energy Calculations: Theory and Applications in Chemistry and Biology *Academic Press*, 2007

[92] Daan Frankel, Berend Smit Understanding Molecular Simulation, Second Edition: From Algorithms to Applications *Springer Series in Chemical Physics*, 2001

[93] E. E. Lattman, K. M. Fiebig, and K. A. Dill. Modeling compact denatured states of proteins. *Biochemistry*, 33(20) 6158-6166, 1994.

[94] Thirumalai D Lee N. Dynamics of collapse of flexible polyelectrolites in poor solvents. *Macromolecules*, 34 3446, 2001.

[95] Haiguang Liu and Yong Duan. Effects of posttranslational modifications on the structure and dynamics of histone h3 n-terminal peptide. *Biophys J*, 94(12) 4579-4585, 2008.

[96] Tak Shing Lo, Boris Khusid, and Joel Koplik. Dynamical clustering of counterions on flexible polyelectrolytes. *Phys Rev Lett*, 100(12) 128301, 2008.

[97] K. Luger, A. W. Mder, R. K. Richmond, D. F. Sargent, and T. J. Richmond. Crystal structure of the nucleosome core particle at 2.8 a resolution. *Nature*, 389(6648) 251-260, 1997.

[98] K. Luger and T. J. Richmond. The histone tails of the nucleosome. *Curr Opin Genet Dev*, 8(2) 140-146, 1998.

[99] Karolin Luger and Jeffrey C Hansen. Nucleosome and chromatin fiber dynamics. *Curr Opin Struct Biol*, 15(2) 188-196, 2005.

[100] Liang Ma and Qiang Cui. The temperature dependence of salt-protein association is sequence specific. *Biochemistry*, 45(48) 14466-14472, 2006.

[101] Raphael Margueron, Patrick Trojer, and Danny Reinberg. The key to development interpreting the histone code? *Curr Opin Genet Dev*, 15(2) 163-176, 2005.

[102] Leonardo Mario-Ramrez, Benjamin Hsu, Andreas D Baxevanis, and David Landsman. The histone database a comprehensive resource for histones and histone fold-containing proteins. *Proteins*, 62(4) 838-842, 2006.

[103] Christopher K Materese, Alexey Savelyev, and Garegin A Papoian. Counterion atmosphere and hydration patterns near a nucleosome core particle. *J Am Chem Soc*, 131(41) 15005-15013, 2009.

[104] Christopher Kroboth Materese, Christa Charisse Goldmon, and Garegin A Papoian. Hierarchical organization of eglin c native state dynamics is shaped by competing direct and water-mediated interactions. *Proc Natl Acad Sci U S A*, 105(31) 10659-10664, 2008.

[105] Yuguang Mu, Phuong H Nguyen, and Gerhard Stock. Energy landscape of a small peptide revealed by dihedral angle principal component analysis. *Proteins*, 58(1) 45-52, 2005.

[106] Samrat Mukhopadhyay, Rajaraman Krishnan, Edward A Lemke, Susan Lindquist, and Ashok A Deniz. A natively unfolded yeast prion monomer adopts an ensemble of collapsed and rapidly fluctuating structures. *Proc Natl Acad Sci U S A*, 104(8) 2649-2654, 2007.

[107] Karl P Nightingale, Laura P O'Neill, and Bryan M Turner. Histone modifications signalling receptors and potential elements of a heritable epigenetic code. *Curr Opin Genet Dev*, 16(2) 125-136, 2006.

[108] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes. Theory of protein folding the energy landscape perspective. *Annu Rev Phys Chem*, 48 545-600, 1997.

[109] Alexey Onufriev, Donald Bashford, and David A Case. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins*, 55(2) 383-394, 2004.

[110] Garegin A Papoian. Proteins with weakly funneled energy landscapes challenge the classical structure-function paradigm. *Proc Natl Acad Sci U S A*, 105(38) 14237-14238, 2008.

[111] Garegin A Papoian and Peter G Wolynes. The physics and bioinformatics of binding and folding-an energy landscape perspective. *Biopolymers*, 68(3) 333-349, 2003.

[112] R. V. Pappu, R. Srinivasan, and G. D. Rose. The flory isolated-pair hypothesis is not valid for polypeptide chains implications for protein folding. *Proc Natl Acad Sci U S A*, 97(23) 12565-12570, 2000.

[113] Rohit V Pappu, Xiaoling Wang, Andreas Vitalis, and Scott L Crick. A polymer physics perspective on driving forces and mechanisms for protein aggregation. *Arch Biochem Biophys*, 469(1) 132-141, Jan 2008.

[114] G. Parisi. Order parameter for spin-glasses. *Phys Rev Lett*, 50 1946-1948, 1983.

[115] Sergei E Permyakov, Anush G Bakunts, Alexander I Denesyuk, Ekaterina L Knyazeva, Vladimir N Uversky, and Eugene A Permyakov. Apo-parvalbumin as an intrinsically disordered protein. *Proteins*, 72(3) 822-836, 2008.

[116] Predrag Radivojac, Lilia M Iakoucheva, Christopher J Oldfield, Zoran Obradovic, Vladimir N Uversky, and A. Keith Dunker. Intrinsic disorder and functional proteomics. *Biophys J*, 92(5) 1439-1456, 2007.

[117] Dirk Reith, Mathias Ptz, and Florian Mller-Plathe. Deriving effective mesoscale potentials from atomistic simulations. *J Comput Chem*, 24(13) 1624-1636, 2003.

[118] J. C. Rice and C. D. Allis. Code of silence. *Nature*, 414(6861) 258-261, 2001.

[119] Daniel R Roe, Asim Okur, Lauren Wickstrom, Viktor Hornak, and Carlos Simmerling. Secondary structure bias in generalized born solvent models com-

parison of conformational ensembles and free energy of solvent polarization from explicit and implicit solvation. *J Phys Chem B*, 111(7) 1846-1857, 2007.

[120] Roger Rousseau, Eduard Schreiner, Axel Kohlmeyer, and Dominik Marx. Temperature-dependent conformational transitions and hydrogen-bond dynamics of the elastin-like apeptide gvg(vpgvg) a molecular-dynamics study. *Biophys J*, 86(3) 1393-1407, 2004.

[121] I. Rouzina and V. A. Bloomfield. Heat capacity effects on the melting of dna. 1. general aspects. *Biophys J*, 77(6) 3242-3251, 1999.

[122] M. Rubinstein and Ralph H. Colby. *Polymer Physics*. Oxford University Press, USA, 2003.

[123] Erik Sandelin. On hydrophobicity and conformational specificity in proteins. *Biophys J*, 86(1 Pt 1) 23-30, Jan 2004.

[124] Alexey Savelyev and Garegin A Papoian. Electrostatic, steric, and hydration interactions favor na(+) condensation around dna compared with k(+). *J Am Chem Soc*, 128(45) 14506-14518, 2006.

[125] Alexey Savelyev and Garegin A Papoian. Inter-dna electrostatics from explicit solvent molecular dynamics simulations. *J Am Chem Soc*, 129(19) 6060-6061, 2007.

[126] Alexey Savelyev and Garegin A Papoian. Polyionic charge density plays a key role in differential recognition of mobile ions by biopolymers. *J Phys Chem B*, 112(30) 9135-9145, 2008.

[127] Joachim Latzer, Garegin A Papoian, Michael C Prentiss, Elizabeth A Komives, and Peter G Wolynes. Induced fit, folding, and recognition of the nf-kappab-nuclear localization signals by ikappabalpha and ikappabbeta. *J Mol Biol*, 367(1):262–274, Mar 2007.

[128] Timothy J Richmond and Curt A Davey. The structure of dna in the nucleosome core. *Nature*, 423(6936):145–150, May 2003.

[129] Thomas Schalch, Sylwia Duda, David F Sargent, and Timothy J Richmond. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature*, 436(7047) 138-141, 2005.

[130] H. Schiessel. The nucleosome a transparent, slippery, sticky and yet stable dna-protein complex. *Eur Phys J E Soft Matter*, 19(3) 251-262, 2006.

[131] Helmut Schiessel. The physics of chromatin. *Journal of Physics Condensed Matter*, 15 R699-R774, 2003.

[132] Eduard Schreiner, Chiara Nicolini, Bjrn Ludolph, Revanur Ravindra, Nikolaj Otte, Axel Kohlmeyer, Roger Rousseau, Roland Winter, and Dominik Marx. Folding and unfolding of an elastinlike oligopeptide "inverse temperature transition," reentrance, and hydrogen-bond dynamics. *Phys Rev Lett*, 92(14) 148101, 2004.

[133] Shantanu Sharma, Feng Ding, and Nikolay V Dokholyan. Multiscale modeling of nucleosome dynamics. *Biophys J*, 92(5) 1457-1470, 2007.

[134] B. A. Shoemaker, J. J. Portman, and P. G. Wolynes. Speeding molecular recognition by using the folding funnel the fly-casting mechanism. *Proc Natl Acad Sci U S A*, 97(16) 8868-8873, 2000.

[135] Michael Shogren-Knaak, Haruhiko Ishii, Jian-Min Sun, Michael J Pazin, James R Davie, and Craig L Peterson. Histone h4-k16 acetylation controls chromatin structure and protein interactions. *Science*, 311(5762) 844-847, 2006.

[136] Bruschweiler Rafael Showalter Scott A.. Validation of molecular dynamics simulations of biomolecules using nmr spin relaxation as benchmarks Application to the amber99sb force field. *Journal of Chemical Theory and Computation*, 3 961- 975, 2007.

[137] A. Kolinski J. Skolnick. Determinants of secondary structure of polypeptide chains interplay between short range and burial interactions. *J Chem Phys*, 107 953-964, 1997.

[138] Andrea Soranno, Renato Longhi, Tommaso Bellini, and Marco Buscaglia. Kinetics of contact formation and end-to-end distance distributions of swollen disordered peptides. *Biophys J*, 96(4) 1515-1528, 2009.

[139] B. D. Strahl and C. D. Allis. The language of covalent histone modifications. *Nature*, 403(6765) 41-45, Jan 2000.

[140] Samantha S Strickler, Alexey V Gribenko, Alexander V Gribenko, Timothy R Keiffer, Jessica Tomlinson, Tracey Reihle, Vakhtang V Loladze, and George I Makhatadze. Protein stability and surface electrostatics a charged relationship. *Biochemistry*, 45(9) 2761-2766, 2006.

[141] Yuji Sugita, Akio Kitao, and Yuko Okamoto. Multidimensional replica-exchange method for free-energy calculations. *The Journal of Chemical Physics*, 113(15) 6042-6051, 2000.

[142] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314(1-2) 141-151, 1999.

[143] Agnes Toth-Petroczy, I. Simon, M. Fuxreiter, and Y. Levy. Disordered tails of homeodomains facilitate dna recognition by providing a trade-off between folding and specific binding. *J. Am. Chem. Soc.*, 131 15084-15085, 2009.

[144] Hoang T Tran, Albert Mao, and Rohit V Pappu. Role of backbone-solvent interactions in determining conformational equilibria of intrinsically disordered proteins. *J Am Chem Soc*, 130(23) 7380-7392, 2008.

[145] Hoang T Tran and Rohit V Pappu. Toward an accurate theoretical framework for describing ensembles for proteins under strongly denaturing conditions. *Biophys J*, 91(5) 1868-1886, 2006.

[146] Vladimir N Uversky and A. Keith Dunker. Biochemistry. controlled chaos. *Science*, 322(5906) 1340-1341, 2008.

[147] Vladimir N Uversky, Christopher J Oldfield, and A. Keith Dunker. Intrinsically disordered proteins in human diseases introducing the d2 concept. *Annu Rev Biophys*, 37 215-246, 2008.

[148] Andreas Vitalis, Xiaoling Wang, and Rohit V Pappu. Quantitative characterization of intrinsic disorder in polyglutamine insights from analysis based on polymer theories. *Biophys J*, 93(6) 1923-1937, 2007.

[149] Andreas Vitalis, Xiaoling Wang, and Rohit V Pappu. Atomistic simulations of the effects of polyglutamine chain length and solvent quality on conformational equilibria and spontaneous homodimerization. *J Mol Biol*, 384(1) 279-297, 2008.

[150] Dana Vuzman, Ariel Azia, and Yaakov Levy. Searching dna via a "monkey bar" mechanism the significance of disordered tails. *J Mol Biol*, 396(3) 674-684, 2010.

[151] Dana Vuzman and Yaakov Levy. Dna search efficiency is modulated by charge composition and distribution in the intrinsically disordered tail. *Proc Natl Acad Sci U S A*, 107(49) 21004-21009, 2010.

[152] Dana Vuzman, Michal Polonsky, and Yaakov Levy. Facilitated dna search by multidomain transcription factors cross talk via a flexible linker. *Biophys J*, 99(4) 1202-1211, 2010.

[153] X. Wang, S. C. Moore, M. Laszckzak, and J. Ausi. Acetylation increases the alpha-helical content of the histone tails of the nucleosome. *J Biol Chem*, 275(45) 35013-35020, 2000.

[154] Lauren Wickstrom, Asim Okur, and Carlos Simmerling. Evaluating the performance of the ff99sb force field based on nmr scalar coupling data. *Biophys J*, 97(3) 853-856, 2009.

[155] J. Widom. Structure, dynamics, and function of chromatin in vitro. *Annu Rev Biophys Biomol Struct*, 27 285-327, 1998.

[156] Roland G. Winkler, Michael Gold, and Peter Reineker. Collapse of polyelectrolyte macromolecules by counterion condensation and ion pair formation A molecular dynamics simulation study. *Phys. Rev. Lett.*, 80(17) 3731-3734, 1998.

[157] Sangwook Wu, Pavel I Zhuravlev, and Garegin A Papoian. High resolution approach to the native state ensemble kinetics and thermodynamics. *Biophys J*, 95(12) 5524-5532, 2008.

[158] Darren Yang and Gaurav Arya. Structure and binding of the h4 histone tail and the effects of lysine 16 acetylation. *Physical Chemistry Chemical Physics*, 13 2911-21, 2011.

[159] Hsiang-Ai Yu and Martin Karplus. A thermodynamic analysis of solvation. *The Journal of Chemical Physics*, 89 2366-2379, 1988.

[160] Chunyang Zheng and Jeffrey J Hayes. Structures and interactions of the core histone tail domains. *Biopolymers*, 68(4) 539-546, 2003.

[161] Chunyang Zheng, Xu Lu, Jeffrey C Hansen, and Jeffrey J Hayes. Salt-dependent intra- and internucleosomal interactions of the h3 tail domain in a model oligonucleosomal array. *J Biol Chem*, 280(39) 33552-33557, 2005.

[162] Pavel I Zhuravlev, Christopher Kroboth Materese, and Garegin A Papoian. Deconstructing the native state energy landscapes, function, and dynamics of globular proteins. *J Phys Chem B*, 113(26) 8800-8812, 2009.

[163] Pavel I Zhuravlev and Garegin A Papoian. Functional versus folding landscapes the same yet different. *Curr Opin Struct Biol*, 20(1) 16-22, 2010.

[164] Pavel I Zhuravlev and Garegin A Papoian. Protein functional landscapes, dynamics, allostery a tortuous path towards a universal theoretical framework. *Q Rev Biophys*, 43(3) 295-332, 2010.

[165] P. Maragakis, M. Karplus Large amplitude conformational change in proteins explored with a plastic network model adenylate kinase. *J. Mol. Biol.*, 352 807, 2005.

[166] Abdollah Allahverdi, Renliang Yang, Nikolay Korolev, Yanping Fan, Curt A Davey, Chuan-Fa Liu, and Lars Nordenskiöld. The effects of histone h4 tail acetylations on cation-induced chromatin folding and self-association. *Nucleic acids research*, 39(5) 1680-91,z 2011.

[167] D Angelov, J Vitolo, V Mutskov, and S Dimitrov. Preferential interaction of the core histone tail domains with linker dna. *Proc. Natl. Acad. Sci. USA*, 98 6955, 2001.

[168] AJ Bannister and T Kozarides. Regulation of chromatin by histone modifications. *Cell Research*, 21 381, 2011.

[169] A G Cherstvy. Positively charged residues in dna-binding domains of structural proteins follow sequence-specific positions of dna phosphate groups. *J. Phys. Chem. B*, 113(13) 4242-7, 2009.

[170] B Dorigo, T Schalch, K Bystricky, and T J Richmond. Chromatin fiber folding requirement for the histone h4 n-terminal tail. *J. Mol. Biol.*, 327 85, 2003.

[171] Alex Gansen, Katalin Tóth, Nathalie Schwarz, and Jörg Langowski. Structural variability of nucleosomes detected by single-pair förster resonance energy transfer histone acetylation, sequence variation, and salt effects. *J. Phys. Chem. B*, 113(9) 2604-13, 2009.

[172] Stephen E Halford and John F Marko. How do site-specific dna-binding proteins find their targets? *Nucleic acids research*, 32(10) 3040-52, 2004.

[173] T Jenuwein and C D Allis. Translating the histone code. *Science*, 293 1074, 2001.

[174] W Jorgensen, J Chandrasekhar, J D Madura, R W Impey, and ML Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79 926, 1983.

[175] In Suk Joung and Thomas E Cheatham. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B*, 112(30) 9020-41, 2008.

[176] Tony Kouzarides. Chromatin modifications and their function. *Cell*, 128(4) 693-705, 2007.

[177] D Lee, J Hayes, D Pruss, and A P Wolffe. A positive role for histone acetylation in transcription factor access to nucleosomal DNA. *Cell*, 72 7384, 1993.

[178] N Lee and D Thirumalai. Dynamics of collapse of flexible polyampholytes. *J. Chem. Phys.*, 113 5126, 2000.

[179] Jianyuan Luo, Muyang Li, Yi Tang, Monika Laszkowska, Robert G Roeder, and Wei Gu. Acetylation of p53 augments its site-specific dna binding both in vitro and in vivo. *Proc. Natl. Acad. Sci. USA*, 101(8) 2259-64, 2004.

[180] V K Misra, J L Hecht, A S Yang, and B Honig. Electrostatic contributions to the binding free energy of the lambdaci repressor to dna. *Biophys. J.*, 75(5) 2262-73, 1998.

[181] V K Misra and B Honig. On the magnitude of the electrostatic contribution to ligand-dna interactions. *Proc. Natl. Acad. Sci. USA*, 92(10) 4691-5, 1995.

[182] F Mühlbacher, C Holm, and H Schiessel. Controlled dna compaction within chromatin the tail-bridging effect. *Europhys. Lett.*, 73 135, 2006.

[183] A Pérez and et al. Refinement of the amber force field for nucleic acids improving the description of alpha/gamma conformers. *Biophys J*, 92 3817, 2007.

[184] Davit A Potoyan and Garegin A Papoian. Energy landscape analyses of disordered histone tails reveal special organization of their conformational dynamics. *J Am Chem Soc*, 133(19) 7405-15, 2011.

[185] Ignacio E Sánchez, Diego U Ferreiro, Mariano Dellarole, and Gonzalo de Prat-Gay. Experimental snapshots of a protein-dna binding landscape. *Proc. Natl. Acad. Sci. USA*, 107(17) 775, 2010.

[186] Mona D Shahbazian and Michael Grunstein. Functions of site-specific histone acetylation and deacetylation. *Ann. Rev. Biochem.*, 76 75, 2007.

[187] R S Spolar and M T Record. Coupling of local folding to site-specific binding of proteins to dna. *Science*, 263(5148) 777-84, 1994.

[188] K Tóth and N Brun.... Chromatin compaction at the mononucleosome level. *Biochemistry*, 45 1591, 2006.

[189] Peter H von Hippel. Biochemistry. completing the view of transcriptional regulation. *Science*, 305(5682) 350-2, 2004.

[190] J Wang, P Cieplak, and P Kollman. How well does a restrained electrostatic potential (resp) model perform in calculating conformational energies of organic and biological molecules? *J. Comp Chem.*, 21 1049, 2000.

[191] Christopher L Woodcock and Rajarshi P Ghosh. Chromatin higher-order structure and dynamics. *Cold Spring Harb Perspect Biol*, 2(5) 1-25, 2010.

[192] Jörgen Adén and Magnus Wolf-Watz. Nmr identification of transient complexes critical to adenylate kinase catalysis. *J Am Chem Soc*, 129(45) 14003-12, 2007.

[193] Bharat V Adkar, Biman Jana, and Biman Bagchi. Role of water in the enzymatic catalysis study of atp + amp 2adp conversion by adenylate kinase. *J Phys Chem A*, 115(16) 3691-7, 2011.

[194] Karunesh Arora and Charles L Brooks. Large-scale allosteric conformational transitions of adenylate kinase appear to involve a population-shift mechanism. *Proc. Natl. Acad. Sci. USA*, 104(47) 18496-501, 2007.

[195] Nilesh K Banavali and Benoît Roux. Free energy landscape of a-dna to b-dna conversion in aqueous solution. *J Am Chem Soc*, 127(18) 6866-6876, 2005.

[196] Michael D Daily, George N Phillips, and Qiang Cui. Many local motions cooperate to produce the adenylate kinase conformational transition. *J. Mol. Biol.*, 400(3) 618-31, 2010.

[197] Paul W Fenimore, Hans Frauenfelder, B H McMahon, and R D Young. Bulk-solvent and hydration-shell fluctuations, similar to alpha- and beta-fluctuations in glasses, control protein motions and functions. *Proc Natl Acad Sci U S A*, 101(40) 14408-14413, 2004.

[198] Hans Frauenfelder, Guo Chen, Joel Berendzen, Paul W Fenimore, Helén Jansson, Benjamin H McMahon, Izabela R Stroe, Jan Swenson, and Robert D Young. A unified model of protein dynamics. *Proc. Natl. Acad. Sci. USA*, 106(13) 5129-34, 2009.

[199] Hans Frauenfelder, Fritz G Parak, and R D Young. Conformational substates in proteins. *Ann. Rev. Biophys. Biochem.*, 17 451-79, Jan 1988.

[200] Hans Frauenfelder, S G Sligar, and Peter G Wolynes. The energy landscapes and motions of proteins. *Science*, 254(5038) 1598-1603, 1991.

[201] G Hammond. A correlation of reaction rates. *J. Am. Chem. Soc.*, 77 334, Jan 1955.

[202] Jeffrey A Hanson, Karl Duderstadt, Lucas P Watkins, Sucharita Bhattacharyya, Jason Brokaw, Jhih-Wei Chu, and Haw Yang. Illuminating the mechanistic roles of enzyme conformational dynamics. *Proc. Natl. Acad. Sci. USA*, 104(46) 18055-60, 2007.

[203] Katherine A Henzler-Wildman, Ming Lei, Vu Thai, S Jordan Kerns, Martin Karplus, and Dorothee Kern. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature*, 450(7171) 913-916, 2007.

[204] Katherine A Henzler-Wildman, Vu Thai, Ming Lei, Maria Ott, Magnus Wolf-Watz, Tim Fenn, Ed Pozharski, Mark A Wilson, Gregory A Petsko, Martin Karplus, Christian G Hübner, and Dorothee Kern. Intrinsic motions along an enzymatic reaction trajectory. *Nature*, 450(7171) 838-44, 2007.

[205] Biman Jana, Bharat V Adkar, Rajib Biswas, and Biman Bagchi. Dynamic coupling between the lid and nmp domain motions in the catalytic conversion of atp and amp to adp by adenylate kinase. *J. Chem. Phys.*, 134(3) 035101, Jan 2011.

[206] Dorothee Kern, Elan Zohar Eisenmesser, and Magnus Wolf-Watz. Enzyme dynamics during catalysis measured by nmr spectroscopy. *Meth Enzymol*, 394 507-24, Jan 2005.

[207] Dorothee Kern and Erik R P Zuiderweg. The role of dynamics in allosteric regulation. *Curr Op Struct Biol*, 13(6) 748-57, 2003. A review of preexisting equilibrium cases.

[208] Peter A Kollman, I Massova, C Reyes, B Kuhn, S Huo, L Chong, M Lee, T Lee, Y Duan, W Wang, O Donini, P Cieplak, J Srinivasan, D A Case, and T E Cheatham. Calculating structures and free energies of complex molecules combining molecular mechanics and continuum models. *Acc Chem Res*, 33(12) 889-97, 2000.

[209] Jan Kubelka, James Hofrichter, and William A Eaton. The protein folding 'speed limit'. *Curr Op Struct Biol*, 14(1) 76-88, 2004.

[210] Yaakov Levy, Samuel S Cho, José Nelson Onuchic, and Peter G Wolynes. A survey of flexible protein binding mechanisms and their transition states using native topology based energy landscapes. *J Mol Biol*, 346(4) 1121-1145, 2005.

[211] Qiang Lu and Jin Wang. Single molecule conformational dynamics of adenylate kinase energy landscape, structural correlations, and transition state ensembles. *J Am Chem Soc*, 130(14) 4772-83, 2008.

[212] A D MacKerell, Nilesh K Banavali, and N Foloppe. Development and current status of the charmm force field for nucleic acids. *Biopolymers*, 56(4) 257-265, 2000.

[213] Dmitrii E Makarov. Computer simulations and theory of protein translocation. *Acc. Chem. Res.*, 42(2) 281-9, 2009.

[214] J M Matthews and A R Fersht. Exploring the energy surface of protein folding by structure-reactivity relationships and engineered proteins observation of hammond behavior for the gross structure of the transition state and anti-hammond behavior for structural elements for unfolding/folding of barnase. *Biochemistry*, 34(20) 6805-14, 1995.

[215] Osamu Miyashita, José Nelson Onuchic, and Peter G Wolynes. Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. *Proc Natl Acad Sci USA*, 100(22) 12570-5, 2003.

[216] C W Müller, G J Schlauderer, J Reinstein, and G E Schulz. Adenylate kinase motions during catalysis an energetic counterweight balancing substrate binding. *Structure*, 4(2) 147-56, 1996.

[217] C W Müller and G E Schulz. Structure of the complex between adenylate kinase from escherichia coli and the inhibitor ap5a refined at 1.9 a resolution. a model for a catalytic transition state. *J. Mol. Biol.*, 224(1) 159-77, 1992.

[218] Ulrika Olsson and Magnus Wolf-Watz. Overlap between folding and functional energy landscapes for adenylate kinase conformational change. *Nature communications*, 1 111, Jan 2010.

[219] J N Onuchic, Z Luthey-Schulten, and Peter G Wolynes. Theory of protein folding the energy landscape perspective. *Annu Rev Phys Chem*, 48 545-600, Jan 1997.

[220] Sanghyun Park, Albert Y Lau, and Benoît Roux. Computing conformational free energy by deactivated morphing. *J. Chem. Phys.*, 129(13) 134102, 2008.

[221] James C Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D Skeel, Laxmikant Kalé, and Klaus Schulten. Scalable molecular dynamics with namd. *J Comput Chem*, 26(16) 1781-1802, 2005.

[222] S Plimpton. Fast parallel algorithms for short-range molecular dynamics. *J Comp Phys*, 117 1, Jan 1995.

[223] Steven S Plotkin, Wang, and Peter G Wolynes. Correlated energy landscape model for finite, random heteropolymers. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics*, 53(6) 6271-6296, 1996.

[224] J Portman and P G Wolynes S Takada. Variational theory for site resolved protein folding free energy surfaces. *Phys. Rev. Lett.*, 23 5237, Jan 1998.

[225] Daniel R Roe, Asim Okur, Lauren Wickstrom, Viktor Hornak, and Carlos Simmerling. Secondary structure bias in generalized born solvent models comparison of conformational ensembles and free energy of solvent polarization from explicit and implicit solvation. *J. Phys. Chem. B*, 111(7) 1846-57, 2007.

[226] Yury E Shapiro, Edith Kahana, Vitali Tugarinov, Zhichun Liang, Jack H Freed, and Eva Meirovitch. Domain flexibility in ligand-free and inhibitor-bound escherichia coli adenylate kinase based on a mode-coupling analysis of 15n spin relaxation. *Biochemistry*, 41(20) 6271-81, 2002.

[227] Yury E Shapiro and Eva Meirovitch. Activation energy of catalysis-related domain motion in e. coli adenylate kinase. *J. Phys. Chem. B*, 110(23) 11519-24, 2006.

[228] S Takada, J J Portman, and P G Wolynes. An elementary mode coupling theory of random heteropolymer dynamics. *Proc. Natl. Acad. Sci. USA*, 94(6) 2318-21, 1997.

[229] Y N Vorobjev and J Hermans. Es/is estimation of conformational free energy by combining dynamics simulations with explicit solvent with an implicit solvent continuum model. *Biophys Chem*, 78(1-2) 195-205, 1999.

[230] Y N Vorobjev and J Hermans. Free energies of protein decoys provide insight into determinants of protein stability. *Protein Sci*, 10(12) 2498-2506, 2001.

[231] Aaron Weis, Kambiz Katebzadeh, Pär Söderhjelm, Ingemar Nilsson, and Ulf Ryde. Ligand affinities predicted with the mm/pbsa method dependence on the simulation method and the force field. *J Med Chem*, 49(22) 6596-606, 2006.

[232] Paul C Whitford, Osamu Miyashita, Yaakov Levy, and José N Onuchic. Conformational transitions of adenylate kinase switching by cracking. *J. Mol. Biol.*, 366(5) 1661-71, 2007.

[233] Magnus Wolf-Watz, Vu Thai, Katherine Henzler-Wildman, Georgia Hadjipavlou, Elan Z Eisenmesser, and Dorothee Kern. Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. *Nat Struct Mol Biol*, 11(10) 945-9, 2004.

[234] Huan-Xiang Zhou. Rate theories for biologists. *Quart. Rev. Biophys.*, 43(2) 219-93, 2010.

[235] Pavel I Zhuravlev, Christopher Kroboth Materese, and Garegin A Papoian. Deconstructing the native state energy landscapes, function, and dynamics of globular proteins. *J. Phys. Chem. B*, 113(26) 8800-12, 2009.

[236] Pavel I Zhuravlev and Garegin A Papoian. Functional versus folding landscapes the same yet different. *Curr Op Struct Biol*, 20(1) 16-20, Jan 2010.

[237] Pavel I Zhuravlev and Garegin A Papoian. Protein functional landscapes, dynamics, allostery a tortuous path towards a universal theoretical framework. *Q Rev Biophys*, 43(3) 295-332, 2010.

[238] Pavel I Zhuravlev, Sangwook Wu, Davit A Potoyan, Michael Rubinstein, and Garegin A Papoian. Computing free energies of protein conformations from explicit solvent simulations. *Methods*, 52(1) 115-21, 2010.