AFM Study of Gene Silencing by DNA Methylation and Its Interactions Involving Chromatin and Methyl CpG Binding Proteins

by

Parminder Kaur

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2012 by the
Graduate Supervisory Committee:

Stuart Lindsay, Chair
Robert Ros
Oliver Beckstein
Sara Vaiana
Nongjian Tao

ARIZONA STATE UNIVERSITY

August 2012

ABSTRACT

CpG methylation is an essential requirement for the normal development of mammals, but aberrant changes in the methylation can lead to tumor progression and cancer. An in-depth understanding of this phenomenon can provide insights into the mechanism of gene repression. We present a study comparing methylated DNA and normal DNA wrt its persistence length and contour length. Although, previous experiments and studies show no difference between the physical properties of the two, the data collected and interpreted here gives a different picture to the methylation phenomena and its effect on gene silencing. The study was extended to the artificially reconstituted chromatin and its interactions with the methyl CpG binding proteins were also probed.

DEDICATION

This thesis is dedicated to my dad. I love you dad. This is for you.

ACKNOWLEDGMENTS

First and foremost I offer my sincerest gratitude to my advisor, Dr Stuart Lindsay, who supported me throughout my thesis with his patience and knowledge whilst giving me freedom to work in my own way. I thank him for providing me with an unflinching encouragement and support in various ways. His clear understanding of the subject combined with his scientific intuition has made him a constant oasis of ideas, which exceptionally inspire and enrich all those who work with him. His support has helped my growth as a student, a researcher and a scientist I want to be. Thank you Dr.Lindsay.

I would like to thank Dr.Hongda Wang, for being the first person who taught me how to work with an AFM. I am proud to record that I had several opportunities to work with an exceptionally experienced scientist like him.

I would also like to thank Dr. Birgit Plochberger, who gave me an immense support and guidance during her stay as a Postdoc in Lindsay lab. I found a great advisor as well a friend in you. Your inputs were invaluable. And your Matlab code rocks . Thanks Birgit.

I would like to thank Peter Costa, with whom I started this project . Even though you left midway, but your input towards this dissertation is indispensible. Being an undergrad, I was always amazed by your hardwork and sincerity towards work. Thanks a lot Pete.

I would also like to thank Qiangfu for designing this DNA template and from whom I learned the basics of biochemistry . You are a great person to work with.

I would also thank Dr. Vaiana and Stephanie Cope for their collaboration in the project in carrying out Dynamic light Scattering measurements. Thanks to both of you for making me understand the basic phenomena and technique of the scattering process.

I would like to thank all the undergraduate and graduate students who worked with me. Trent Bowen , your dedication towards your work is remarkable. I thank you for all the reconstitution procedures and chromatin extraction procedures and many other things you have done for me. Ian Vicino, I thank you for all the DNA you have extracted from those gels. I know it was pretty monotonous, but, you still did it happily and pretty nicely. David Dotson, I just loved

your technique, the way you observed everything. Thanks for all the work you did in the lab for this project even though I still cannot figure out why the bacteria was angry at you and not growing when it would grow for anyone else. I would like to thank Karan Syal and Amber Strunk for all the intelligent discussions I had with them . Specially Karan ,for your suggestion of DMSO as a solvent for the experiment.

I would also like to thank all my fellow students and post docs in Lindsay lab.

My special thanks to Mike Dodson,, without whom I would never have a got a single sample delivered. I know I owe you a semitruck of Pina coladas and you will not have to remind me of it. I will surely get it for you and you can have it by your favorite place, poolside. Thanks Mikeeeee.

I would also like to thank my lab manager Steve Woodward.

I convey special acknowledgement to Margaret Black,for her indispensable help, support and love. You are a motherly figure to me and I love you a lot.

I would like to thank my roommate and friend, Anindita Dey, who actually helped in my lab procedures being a computer engineer. Thanks for all the nights you have stayed awake with me trying to figure to out where exactly I was going wrong in my experiments and taking those AFM images.. I thankyou for helping me out in all those Matlab codes which were greek to me initially. I would like to thankyou for your constant support all through out. Thanks Tinni.

Where would I be without my family? My parents deserve special mention for their inseparable support and prayers. My Father, Dhanwant Singh, in the first place is the person who inspired me to do my doctorate. My Mother, Beant Kaur, is the one who sincerely raised me with her caring and gentle love. I am what I am today because of both of you. Apinder and Manvinder, thanks for being supportive and caring siblings. Jagdeep Singh, my bro-in-law, for constantly supporting me to write my thesis and scaring me by reminding me of the days left to defend. Seriously you were pretty good at it. Manpriya, my sis-in-law,for all the support she gave me.And my nephew and nieces, Ashmeet , Eileen and Sahej, whose one smile can get me out of the deepest mess I am in. Thanks to all three of you.

I would like to thank everybody who was important to the successful realization of thesis.

And finally, I would like to thank God for being with me throughout and showing me the correct path.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

FIGURE Page

FIGURE Page

TERMINOLOGIES

AFM	Atomic Force Microscope
STM	Scanning Tunneling Microscope
SEM	Scanning Electron Microscopy
TEM	Transmission Electron Microscopy
EM	Electron Microscopy
DNA	Deoxyribonucleic Acid
m-DNA	Methylated DNA
m-Chromatin	Methylated Chromatin
TREC	Simultaneous Topography And Recognition Imaging
AAC	Acoustic AC mode
CpG	Cytosine phosphodiester Guanine
DC	Digital Circuit
MAC	Magnetic AC mode
PEG	Polyethylene Glycol
pN	PicoNewton
nN	NanoNewton
Si	Silicon
Si_3N	Silicon Nitride
UV	Ultra Violet
SiOH	Silicon Hydroxide
APTES	Aminopropyltriethoxysilane
NHS	N-Hydroxysuccinimide
MAL	Maleimide
SATP	N-succinimidyl-S-acetylthiopropionate

NH_2OH	Hydroxylamine
HOPG	Highly Ordered Pyrolytic Graphite
K	Potassium
GD	Glutaraldehyde
VE	Vascular Endothelial
HBSS	Hank Buffered Salt Solution
Ca	Calcium
HUVEC	Human Umbilical Vein Endothelial Cells
VEGFR	Vascular Endothelial Growth Factor Receptor
ATP	Adenosine Triphosphate
QD	Quantum Dot
RNA	Ribonucleic acid
HDAC	Histone Deacetylase
SWI/SNF	Switch/Sucrose NonFermentable
HTF	HpaII Tiny Fragments
CpA	Cytosine phosphodiester Adenosine
TpG	Thymine phosphodiester Guanine
HPRT	Hypoxanthine phosphoribosy transferase
l_p	Persistence Length
FJC	Freely Jointed Chain
WLC	Worm Like Chain
LB	Lysogeny Buffer
SAM	S-Adenosine Methionine
Htert	Human Telomerase Reverse Transcriptase
DMSO	Dimethyl sulfoxide
MBD	Methyl Binding Domain
FRET	Fluorescence Resonance Energy Transfer

TE	Tris EDTA
DTT	Dithiothreitol
NaCl	Sodium Chloride
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
EDTA	Ethylenediaminetetraacetic acid

INTRODUCTION

In the current world of biotechnologies, Atomic force microscopy serves as one of the most prominent tools, which can not only help in finding about the topographic properties of the biomolecules, but also can tell about their interactions with their cognate surroundings. The instrument, originally developed in the late eighties, by Binnig, Quate and Gerber [1], for imaging hard surfaces to atomic resolution, has been playing one of the burgeoning role in the field of molecular biology [2]-[5], nanomanipulation [6]-[7] and nanoassembly [8]. It was developed to overcome the basic drawback of STM [9] of imaging only conducting or semiconducting surfaces.

The first applications of AFM were found in the field of material science. Till date, the field of material science is making full use of this technology in the field of improved photovoltaic cells [10], semiconductors [11], thin films [12] and crystallization processes [13]. Nanotechnology is also one of the field were the AFM has been a driving force. AFM cannot only manipulate the materials but it can also provide their images at a reasonable signal to noise ratio. The application of the AFM in biological world was revolutionized with the introduction of the vibrating mode of AFM in air [14] and further in the liquid, known as the tapping mode [15]. In the biological world, with the newly developed imaging techniques and instrumentation, AFM is emerging as a great companion to most of the microscopy techniques like fluorescence spectroscopy [16], Electron microscopy (EM)[17] and X-ray crystallography [18] .The advent of the technique of TREC (Simultaneous topography and recognition)[19], [20] coupled with force spectroscopy [21], [22], has given a great depth to the unknown world of life sciences.

The emergence of the technology of TREC has opened new possibilities for the measurement of inter- and intra-molecular forces [23] at the single molecule level along with detecting specific interactions between the single ligand –receptor pairs [24] simultaneously with a high-resolution topographic image at nm resolution. It offers the smallest force sensor, which can therefore, provide the highest lateral resolution. It can predict very crucial information about the reaction kinetics, structural information of the binding pockets and molecular dynamics for the complexes [25], [26] formed during the biomolecular processes which can give deeper insights into the understanding of receptor ligand recognition. With these important applications, TREC has

been an important step towards understanding the human immune systems which plays a crucial role in understanding various fatal diseases like cancer, HIV /AIDS , hepatitis and the list goes on.

This thesis primarily deals with applying the technique of AFM to get critical information about gene silencing and expression due to the effect of DNA methylation. This includes the following projects:

Project 1: Surfactant action of the methylated DNA in silencing gene expression

Project 2: Study of the physical differences in methylated and normal chromatin and their interactions with methyl CpG binding proteins.

Project 3: Identification of methyl CpG binding proteins in the chromatin extracted from rat brain cells.

Chapter 1

INTRODUCTION TO ATOMIC FORCE MICROSCOPY

The ability of the AFM to image in vacuum, air or in liquids, with nm sized details, obtain high resolution structural details of the biological samples, manipulate samples with nm scale features and to measure forces with more than pico-Newton resolution, makes it a driving technology in the current scientific research as compared to the most microscopy techniques (Table 1) being used. Most importantly, the samples can be imaged in their native state without undergoing any changes to their states.

	AFM	**TEM**	**SEM**	**OPTICAL**
Max. Res.	Atomic	Atomic	1's nm	100's nm
Scan Range	up to 80um	5nm- 500um	50nm -1cm	> 1um
Typical Cost (K$)	100 – 200	≥ 500	200 – 400	10 – 50
Imaging Environment	Air, fluid, vacuum, special gas	Vacuum	Vacuum	Air, fluid
In-Situ	Yes	No	No	Yes
In- fluid	Yes	No	No	Yes

Table 1 Comparison of AFM and Other Microscopy Techniques [27]

The purpose of using an AFM is to characterize the sample by bringing the probe near to the sample surface (0.2 -10nm) .The probe, with a radius of few nm, on the end of the cantilever, is the most important component of the AFM. It can be considered as a spring. The force acting between the probe and the surface is dependent on the spring constant of the probe and the distance between the probe and the sample surface. Hooke's law as gives this:

$F= -k.x$ Where F: force acting

k: spring constant

x: deflection

For the cantilever to sense the deflection, the spring constant of the cantilever should be lower than that of the surface. Attractive or repulsive forces result due to the interaction between the tip and the sample and cause the positive or negative bending of the cantilever. These forces cause a deflection of the cantilever on which the tip is mounted. These changes are monitored using an optical lever detection system, which amplifies cantilever deflections as explained in Figure 1B.

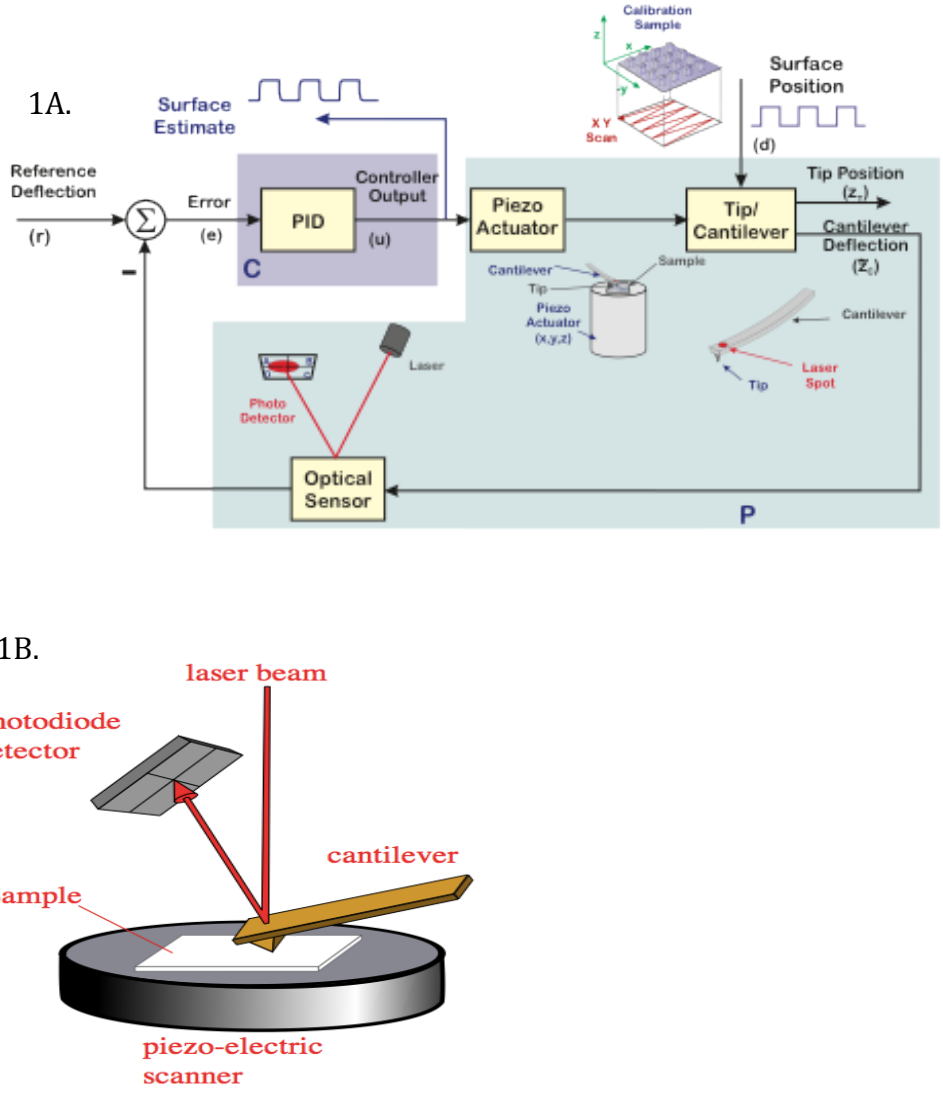


Figure 1 A. An AFM control Block Diagram [27] B. Basic working of an AFM

The laser beam is bounced off of the back of the gold-coated cantilever, which is sensed by the position sensitive photo detector, and signal is given out. There are two modes of operation of AFM:

1. Contact Mode
2. Tapping Mode

The modes of AFM are strictly related to the region of the force field between the tip and the sample as depicted in figure 2.

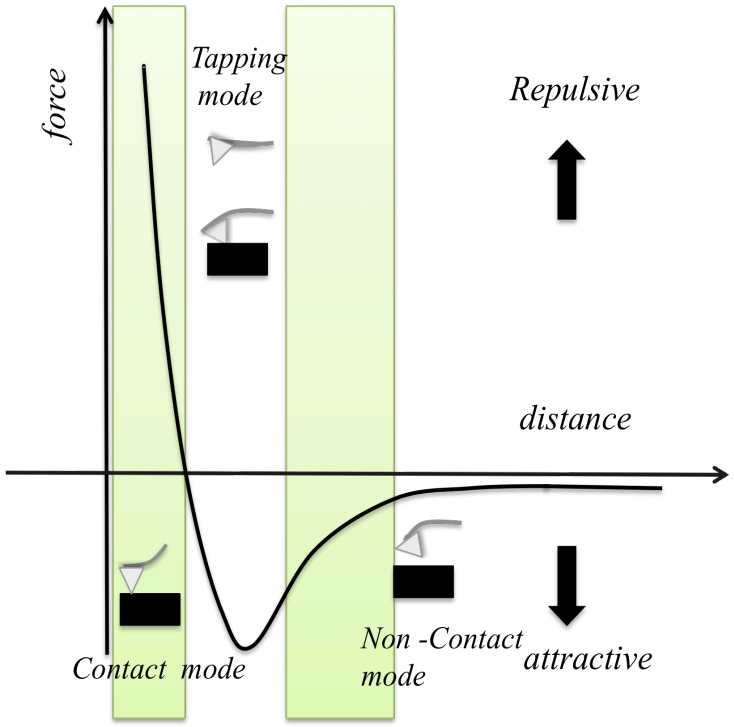


Figure2. Interatomic force variation Vs distance between AFM tip and sample. Repulsive force is the short-range force (< 1nm) and the attractive is the long range one (Vander Waal's force < 100 nm).

1.1 MODES OF AFM

1.1.1 CONTACT MODE AFM

This is basic mode of operation of AFM and is very widely used. Here the tip raster scans in a very close contact with the sample surface, typically less than 0.5nm, and is mainly in the

repulsive regime (refer Figure 2.). With reference to Figure 1A, the interaction forces (repulsive with a mean value of nN) between the tip and the surface cause the cantilever to bend to accommodate changes to topography. The deflection is sensed by the photo detector through the movement of the laser and is compared to the reference deflection in the DC feedback amplifier. If the measured deflection is different from the reference deflection, the feedback amplifier in the controller tries to maintain the desired value, which is a constant value of the deflection, by applying a voltage to the piezo, which raises or lowers the sample relatively to the cantilever. This applied voltage to the piezo is the measure of the height of the sample.

The nonlinear tip sample interaction described above can be explained by Lennard Jones Potential (refer figure 2). At the right side of the curve, the atoms of the tip and the surface are at large distances. When the tip approaches the surface and comes nearby, the atoms first weakly attract each other. This attraction keeps on increasing until the atoms are so close that their electron cloud starts to repel each other. This repulsive force due to the electron clouds overpowers the attractive force as the Interatomic separation decreases and tends to make the force to zero values when the atoms are close enough for the length of a chemical bond. When the Vander Waals forces between the atoms become fully repulsive the atoms are in contact. As can seen from figure 2, the slope of the repulsive regime is very steep, making it balance any force that attempts to push the atoms close together.

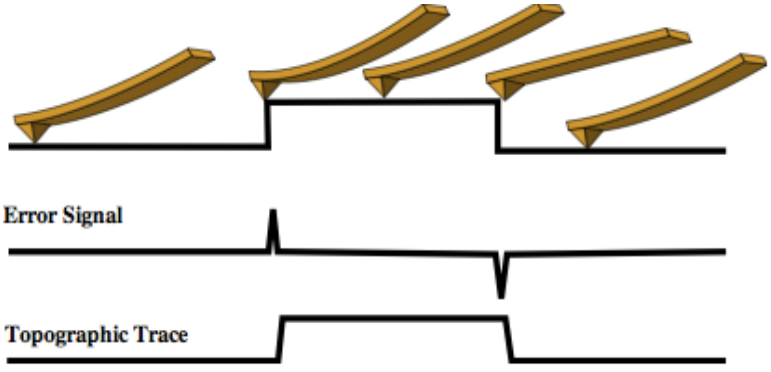


Figure3. Schematic of the signal generation in Contact Mode AFM [28]

In addition to this repulsive Vander Waals force, there are two additional forces in the contact AFM: capillary force due to the presence of the contaminants over the sample surface, which remains attractive throughout as well as the force exerted by the cantilever itself which depends on the deflection and spring constant of the cantilever.

Contact mode is good for high resolution imaging of flat samples such as 2- dimensional arrays of proteins or DNA, where lateral forces are small because of the small variability in the height.

To minimize the force applied on the sample, the spring constant of the cantilevers is taken to be low, in the range of 0.2 – 0.01 N/m. The potential resolution is higher and the scan speed is also higher. However, as significant force must be applied to overcome the effects of surface contamination, soft samples get destroyed during scanning due to the development of the drag force as the tip is scanned across the surface.

1.1.2 TAPPING MODE

To overcome the problem of working with the biological samples, another mode of operation, namely, tapping mode [14], [15] was developed. The tip on the cantilever hovers around 50- 150 Å above the sample surface. The interaction force acting in this case is mostly attractive and thus the problems that arose in the contact mode due to the adhesion and frictional forces can be easily overcome.

In this mode, the tip is oscillated at or near its resonance frequency, and the amplitude of the oscillation is monitored as the tip approaches the surface. It is the scanner that applies force at the base of the tip holder and causes the tip to vibrate. The tip is oscillated at amplitude of range 20-100nm when it is not in contact with the surface. The tip slowly moves towards the surface until it slightly starts to tap the surface. During the scanning process, the tip intermittently touches the surface of the sample and lifts off. The surface is treated as unknown deflection disturbance input .The oscillation amplitude is kept constant through a feedback loop wherein any change to the oscillation amplitude is taken as an error signal relative to the constant reference amplitude... When the tip passes over a bump in the surface, its vibration amplitude is bound to decrease due to

the vibrating space and when it passes over a depression, its vibration amplitude increases (approaching its free air amplitude). Both the times the change in the oscillation amplitude to the reference amplitude is taken as the error signal ,which is sent to the feedback loop in the controller and is given as the topographic features of the sample on the substrate.[5],[26]

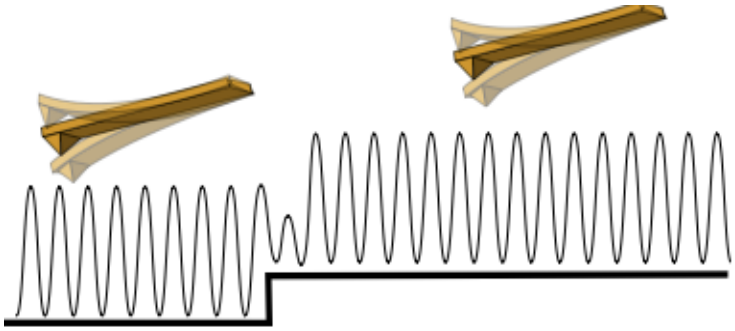


Figure 4 .Tapping Mode AFM [28]

Tapping mode in air must be carried out with stiff cantilevers to avoid the jump in contact problem due to the layer of the adsorbed water and contaminants giving rise to capillary forces and adhesion. This mode in liquid gives the most reliable images as it overcomes the problems due to the contamination layer on the surface. Also, with the liquid environment, it provides for the type of in vivo environment, required for the study of the biological samples. But, as the cantilever has to displace a large amount of solvent as it scans across the surface while oscillating, there are significant viscous losses. Also, it leads to an effective increase in the mass of the cantilever, which leads to lower resonant frequency. Overall it becomes difficult to drive the cantilevers.

1.1.3 MAC MODE AFM

This technique was introduced [30] as it is very gentle and nondestructive for the biological samples. It has the best possible control available for the oscillating probe technology and thus it gives an edge to the researchers for the characterization of the soft samples in liquids with high resolution and sensitivity. It utilizes the magnetic field generated by the solenoid under the sample plate holder, rather than the scanner, for driving the tip. Thus, the rest of the equipment is left unperturbed, giving a very clean cantilever response without any artifacts or background

signals, which were a problem in AC tapping mode. As in the tapping mode, here also cantilever is oscillated at the resonant frequency and the surface topography is given out as a result in the changes in the amplitude of the tip, as it scans across the surface. As, it's the cantilever that is oscillated directly using the magnetic field and there is no need for the scanner vibration, the background resonance is completely removed. This gives increased signal to noise ratio and thus, facilitates the use of cantilevers with lower spring constant, which do lesser damage to the sample and preserve the asperity of the probe, giving increased resolution to the images. As a result, MAC mode AFM is the most used mode for imaging biological samples. [31]

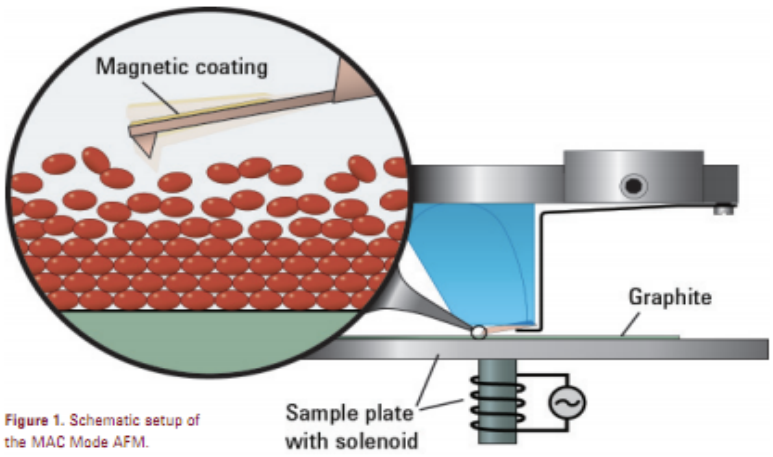


Figure 5 MAC mode AFM [29]

MAC Mode AFM consists of driving a magnetic material coated cantilever by using the magnetic field produced by a solenoid under the sample plate as shown in Figure 5. As there is no perturbation to the microscope directly, the energy that is applied to the sample is very much reduced. This increases the signal to noise ratio. The most sensitive arrangement for the production of the magnetic field is in the direction perpendicular to the moment, M, on the cantilever, resulting in a torque, N= M x B, or a force on the end of the cantilever of magnitude N/L where L is the length of the cantilever. The magnetic field produced by the copper coiled solenoid is given by:

$$B = \frac{\mu_0 NIR^2}{2(z^2 + R^2)^{3/2}} \quad \text{..... [1]}$$

Where B is the vertical component of the magnetic field, μ_o is the permeability of free space, R is the radius of the copper coil, N is the number of turns, I is the current flowing through the coil and z is the distance away from the center of the coil along the symmetry axis. [32]

As the force acting on the sample as well the probe is reduced a lot, this preserves the asperities on the probe, which contributes to enhanced resolution of the order of a nm, absent background resonance and improved signal- to-noise.

1.2 FUNDAMENTAL ELEMENTS OF THE ATOMIC FORCE MICROSCOPE

1.2.1 CANTILEVERS AND TIPS

The tip is the most sensitive and important part of the AFM. These were first handcrafted by using thin films of Gold (Au) foil [1] and tungsten wires [33] respectively. Tips are located at one end of the cantilever as seen in the SEM image in Figure 6. [34]. These are designed differently according to different modes of AFM. They are mostly made up of Silicon and Silicon nitride. Silicon is one of the basic materials, but not the ideal material for the cantilevers, as it wears off easily while imaging hard surfaces. Also, it is reactive and tends to oxidize. Silicon nitride is a much better choice as it is an inert material, and, also much harder than silicon.

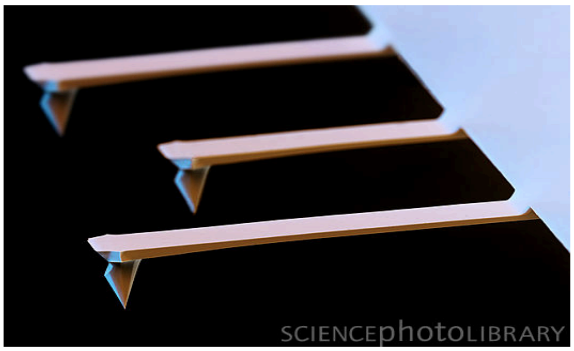


Figure 6 SEM image of Si tip

The fundamental mechanical characteristics of the AFM cantilevers are:

1. Spring constant
2. Resonance frequency
3. Physical dimensions

parameter	Value range	Contact	AC/MAC	AC
Length (μm)	90-460	450	225	125
Width (μm)	25-60	50	28	30
Thickness(μm)	0.7-7.5	2	3	4
Force constant k(N/m)	0.01-91	<0.2	0.1-3	40
Resonant frequency(kHz)	7-420	20	75	300

Table 2 Value Range for cantilever characteristics

The optimal value of the spring constant not only depend on the sample being imaged, but also on the mode of operation. Contact mode cantilevers are much more flexible as compared to the tapping mode cantilevers because of the tips being in constant contact of the surface. With the higher value of spring constant, there is an increase in the amount of the force exerted on the sample, which increases the risk of scratching or destroying the sample. In the noncontact mode, as the cantilever hovers at a distance from the sample, higher spring constant is preferential so that the Vander Waals forces of attraction don't snap the cantilever into the surface. For the contact mode as the tip is touching the surface intermittently, to overcome the hydrodynamic force of attraction, the spring constant of the cantilever is preferred to be high. [35], [36]

The resonant frequency(ν_r) is the frequency of the first natural mode of vibration of a cantilever. If the ν_r of the cantilever is very low, it will couple the ambient frequencies. Also, higher the ν_r, the faster the tip can respond while imaging and thereby does not limits the speed of

response of the surface topography. Also, the cantilevers must have ν_r lower than the bandwidth of the AFM electronics, which is typically around 500kHz.

As the cantilever is being used further in the optical deflection system to reflect the laser, the physical dimensions of the cantilevers are also one of the important parameters. The length of the cantilever should be long enough so that the laser spot can be focused at the end of the cantilever and should provide enough clearance to approach the sample. But, the major constraint on the length is that the sensitivity of the optical deflection system decreases as the length increases. This is because, when the optical deflection system measures the angle of the bend, a longer cantilever bends at a smaller angle as compared to the shorter cantilever for the same deflection. The radius of the tip gives the lateral resolution of the AFM on flat surfaces. The dull tips are sometimes preferable over sharp tips when imaging very soft samples to avoid poking through and destroying the samples. The radius of the tip should be much smaller than the radius of the corrugations in the sample; otherwise it can lead to defective reading of the sample.

The height of the tip determines the ability to characterize a deep structure. The higher the height of the tip, the deeper the structure that can be probed. But, the maximum depth that can be achieved is more depended on the maximum vertical extension of the scanner. Thus, there is not much advantage in making a tip longer than 10-12 um. [35], [36]

1.2.2 SCANNER

AFM scanners are made from piezoelectric material, which have property of expansion and compaction on application of voltage. Polarity of the applied voltage decides whether the scanners will elongate or contract. It is the construction of the scanner by independent combination of the piezo electrodes for X, Y and Z into a single tube that makes it possible for the AFM to take 3D images with high precession.

1.2.3 PHOTODIODE

Figure 7 shows the optical lever deflection of cantilever deflection. The photodiode is divided into four quadrants, A, B, C and D. The amplitude signal is given by the response of the

cantilever deflection in the vertical direction and the frictional signal is given as the response of the cantilever in the direction parallel to the scan. They are respectively given as e^{nor} and f^{nor}. [27]

$$e^{nor} = \frac{(A+D)-(B+C)}{A+B+C+D}$$

$$f^{nor} = \frac{(A+B)-(C+C)}{A+B+C+D}$$

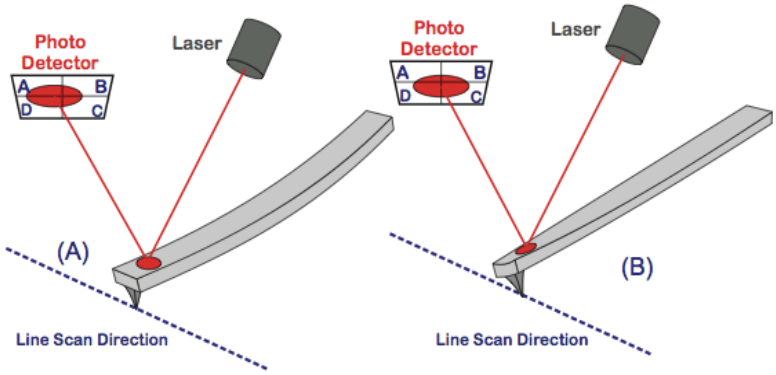


Figure 7 Optical lever detection of cantilever deflection. (A) showing Deflection (B) showing first torsional mode i.e. friction [27]

1.3 RECOGNITION IMAGING

Molecular recognition is one of the most basic aspects in the understanding of the most complex life processes, involving replication, metabolism and signal processing. This process, which relies on the specific behavior among the molecules among themselves, not only helps in studying reaction pathways, enzymatic activities, secretions, genomic replication, transcription and most importantly the responses of the immune system, but also plays an underlying role in the drug design industry to combat the most dreadful diseases occurring on this planet.

Recognition imaging is the technique developed to probe the molecular interactions, for the better understanding of the biomolecular processes at the single molecule level. The main focus of these studies lies in the interaction between receptors and their cognitive ligands, which can be viewed as a paradigm for molecular recognition. It involves converting the AFM tip into a single biomolecular sensor, probing the interactions and molecular dynamics of the receptor ligand

pairs at their typical timescales. Basically, the process involves a force being exerted on a receptor ligand complex and the dissociation process followed over a period of time.

The process involves a technique, termed as TREC [19], [20], giving out a high resolution topographic image overlapped with the recognition image over nm scale resolution. It utilizes the MAC mode AFM [29]-[32], in which the tip is oscillated near its resonance frequency, using the alternating magnetic field produced by the current flowing through the solenoid in the sample plate. The tip is functionalized with a biomolecule [37],, at the end of an elastic polyethylene glycol tether (PEG) [38], [39], and the binding of the biomolecule to its surface bound receptor is studied by applying force to reduce the lifetime of the complex formed until the bond, involving hydrogen bonding and other weaker short range interactions, breaks at a measurable rupture force.

Since the development of this technique in the early 90's, it has not only given insights into the receptor ligand pair interactions[44]-[46], but also has extended its reach to the area of chromatin structures[19],[40]-[43], protein lattices [47]and most lately, the cells[48,49,50].

1.3.1 FUNCTIONALIZATION OF THE CANTILEVERS

Different methodologies have been used over years for the functionalizing of the AFM tips for recognition studies. As the tips (made of Si_3N_4 made with a superficial oxide layer) easily allow for strong attachment of proteins via physisorption, it is one of the earliest and easiest mechanisms for the attachment of proteins. The most common model used for this methodology is biotin- avidin interaction. The main drawback for ligand attachment by this method is that the adsorption and biotin-avidin interactions [~ 50pN] will not be always stronger than most of the protein-ligand interactions. It is for this reason that the covalent interactions are the most reliable methods for functionalization of the cantilevers. [51]

As discussed, the biomolecules should be tightly attached to both the probe and the surface. For very obvious reasons, the interaction between receptor ligand should be weaker as compared to their respective interactions to the probe and the surface. This is usually ensured by using covalent linkages (> 1nN), which are almost 10 times stronger than the typical receptor-ligand linkages (0.2nN) at around the same loading rates. [51]The strong attachment of the

biomolecules to the tips is done using PEG tethers. This not only gives strong attachment, but also provides free diffusion and motion of the ligand about the tip for unconstrained attachment or recognition to the receptor on the surface. It also helps to increase the binding probability significantly. [38], [52]

As discussed in section 1.2.1, AFM cantilevers are made from Si or Si_3N. During the manufacturing and shipping processes, they can pick up surface contaminants, which should be removed so that they don't interfere in the coupling processes. This is done by first rinsing them in ethanol and then further subjecting them to the UV ozone treatment for 15 min. This not only cleans the probes but also introduces a continuous layer of SiOH groups on the surface. [37]
Further, amine groups are generated covalently at a low surface density on the tip apex by reacting the UV cleaned MAC levers. Two most common methods for the amine functionalization are using:

1. Ethanolamine [53] - [55]

2. Vapor deposition of Aminopropyltriethoxysilane. [56]- [59]

All along this study, the vapor deposition method was used as it is considered to be better than the ethanolamine deposition. This step is the most crucial step of the process, as the density of the amine groups specifically determines the number of the ligands on the tip for the recognition process. As these experiments require single molecule interactions, these numbers have to be very low. Typically, these densities are in the range of 200-500 molecules per square micron. Considering the tip radius to be around 5-20 nm, this approximates to around one molecule per effective tip area. During this process, the amine groups present in APTES, need to get hydrolyzed. But, aminosilanes have a tendency to get autohydrolysed and thereby lead to APTES condensing into networks. To prevent premature hydrolysis and condensation, a base in the form of N, N diisopropylethylamine is added during the reaction for lowering the pH to maintain basic conditions, so that the self-hydrolysis of the aminosilanes is prevented. Rigorous exclusion of water also helps the process and this is the reason for the inert atmospheric conditions for the functionalization. The process involves the addition of 30 μl of APTES

(Aminopropyltriethoxy silane)) and 10 μl of the N, N diisopropylethylamine to a desiccator containing the oxidized AFM MAC levers, being purged with Argon. The incubation time for this reaction is anywhere between 40- 60 min.

The next step involves the attachment of a heterobifunctional [59] (NHS-PEG- MAL) flexible and distensible cross linker [38], [39] to the amine groups on the tips. PEG tether is a water-soluble, nontoxic and nonadhesive polymer. Due to these properties of this linker, nonspecific tip –probe interactions are practically absent. Also, these tethers allow an unbound ligand to freely orient and diffuse within a certain volume imparting additional degrees of freedom that are necessary for the ligand to freely orient and bind to the binding site. The tips are added to a solution of the linker in 500μl chloroform and 5μl of triethylamine. Triethylamine, being an aprotic solvent, helps to slow down the hydrolysis of the NHS ester. The reaction is allowed to stand for 2-3 hrs.

The next step involves the attachment of the ligand to the other end (Maleimide) of the PEG linker. Maleimide group being thiol reactive binds to the modified ligand sulphydryl groups. These groups not being present in ligand in the native state are introduced by the reaction of the ligand with N-succinimidyl-3-(S-acetylthio) propionate (SATP). The thiol groups are activated for binding with the maleimide end of the cross linker using the NH_2OH reagent. Ligand functionalized tips are stored in the PBS buffer at 4^0C and they retain their functionality for approximately 10 days.

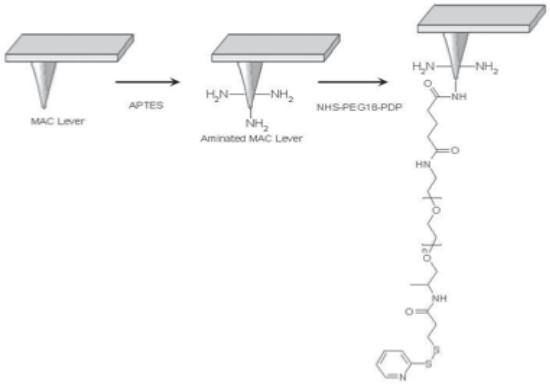


Figure 8 Process of Amination and Pegylation of the MAC lever. [37]

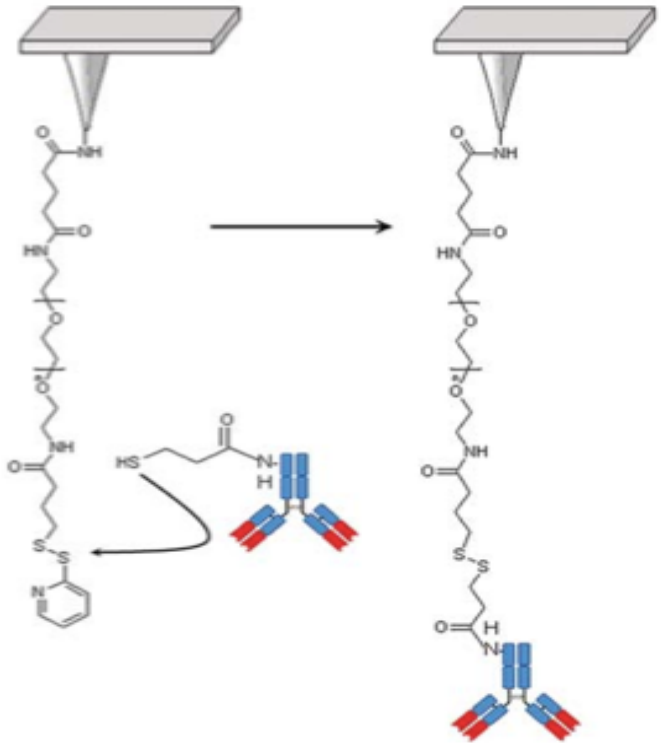


Figure 9 Conjugation of SATP modified Antibody to the Pegylated MAC lever. [3]

1.3.2 SURFACE FOR MOLECULAR RECOGNITION FORCE MICROSCOPY

For the recognition studies to take place, particular surface chemistry needs to be done to allow for tight attachment of the samples to the surface and to avoid nonspecific tip sample interactions. If the study is being done on cells, organelles and big molecules, the flatness of the substrate is not that critical parameter. But, in the case of proteins, the flatness has to be at nm scale as the dimensions of the proteins are also in the same range. [46], [61] The most planar surfaces used for these studies were Highly oriented pyrolytic graphite (HOPG) and mica. The cleavage of these surfaces gives both, ultra flat and clean surfaces. But, HOPG has not been used much except for imaging protein crystals in AFM studies [62].

Muscovite Mica $[(KF)_2(Al_2O_3)_3(SiO_2)_6(H_2O)]$ is the most widespread used surface for AFM samples. It is a phyllosillicate, very hydrophilic and exhibits negative surface potential. This negative potential can be utilized in attaching the proteins, which are cationic in nature, at physiological pH using electrostatic interaction forces. Negatively charged samples can be

immobilized on the mica surface by using a solution of bivalent cations. In the mica surface, K^+ ions are attached electrostatically on the SiO_2 layer. Due to the electrostatic interaction, there can be an exchange of K^+ for Mg^{++}, which can be exploited for the electrostatic attachment of DNA and proteins (anionic in nature). Along with using cationic solutions, a thin layer of poly-L-Lysine can also be used as it changes the surface potential into anionic form. [63]-[65], [21],[51]The nonspecific electrostatic attachment of samples, though, is a very easy technique for the attachment of the biomolecules , but it is very limited and it offers no means to orient the molecules in a particular orientation .

The solution in these cases is the covalent immobilization. For covalent immobilization, amine groups have to be attached on the mica surface. The density of the amine groups depends on the SiOH groups. This can be increased by water plasma treatment. [66]. This process is done by using APTES as already explained in section 1.3.1. After the mica substrates are APTES modified, 50µl of 2µM Glutaraldehyde (GD) is pipetted onto the APTES mica surface. After 12 min of the incubation time, GD mica is rinsed with the DI water and the sample is added on to it and incubated from 40-60 min. [56] .This process is useful only if the samples to be imaged have free lysine residues.

One of the most recent and useful application of the AFM is its ability to image and analyze the living cells. .The sample prep for the same is a very intricate process. Most of the cells are grown in culture on polystyrene or coated (collagen, entactin) plates. [48]- [50]

1.3.3 PRINCIPLES OF TOPOGRAPHY AND RECOGNITION IMAGING

TREC imaging uses MAC mode AFM, which involves driving the magnetized cantilevers by an alternating magnetic field, which is generated by a solenoid under the sample plate holder. The magnetic tip, which is being driven close to its resonance frequency due to the magnetic field, is in intermittent contact with the surface in the downward motion of the tip. This reduces the contact time of the interaction between the tip and the sample, further reducing friction. [29]-[32]While scanning across the surface, tip deflects wrt the topography of the sample, thus giving the topo of the sample .The cantilever , oscillates due to the effect of the magnetic field in the z direction

according to the preset frequency and the peak to peak amplitude. The maxima and the minima of this amplitude, change according to the interaction between the tip and the sample. During the scan, if the deflection signal is seen over a compressed time axis, the changes in the minima of the amplitude clearly gives the topography of the sample .This can be seen in figure 11.

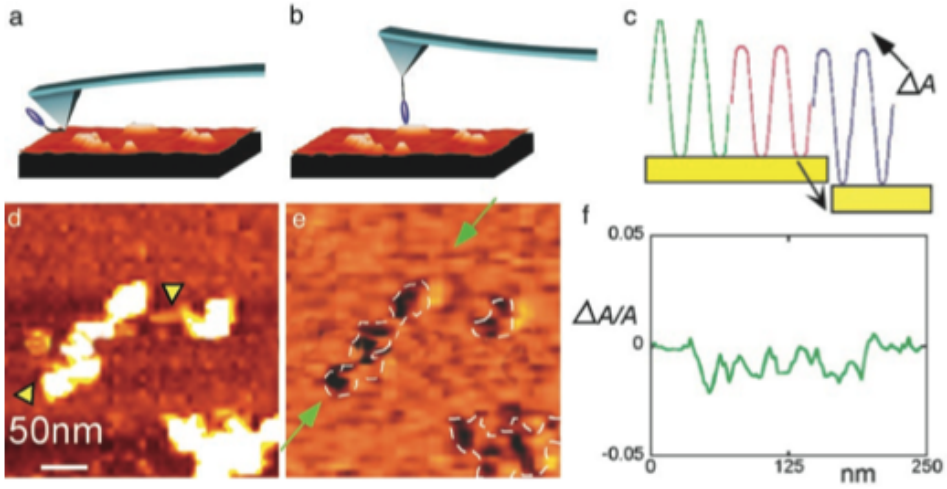


Figure 10. Simultaneous Topographic and Recognition [19]. When an AFM tip-tethered antibody (blue blob in a) binds to its antigen in the sample being scanned (b), there is a transient reduction in the oscillation amplitude of the tip (green curve to red curve in c). The imaging servo restores the signal amplitude but with the peak signal shifted downward by an amount ΔA (blue curve in c). This peak shift provides the recognition signal for a specific antigen–antibody recognition event. A topographic image of MMTV arrays and the corresponding recognition image (a ‘‘map’’ of the change in peak signal, ΔA) are shown in d and e, respectively. These images were obtained simultaneously from one scan of the sample by using an H3-specific antibody tethered to the AFM tip. A plot of the peak signal, ΔA, for the portion of the recognition image between the green arrows (e) is shown in f.

The maxima of the amplitude signal, remains constant only to be effected by the thermal noise of the cantilever. This is the case for bare tip imaging of the samples. But, when the tip is functionalized with an antibody or protein of interest according to the Functionalization procedure

in section 1.3.1, the response is quite different and gives very valuable information about the interactions going on between the protein samples on the tip and the substrate.[20], [61], [67], [68]

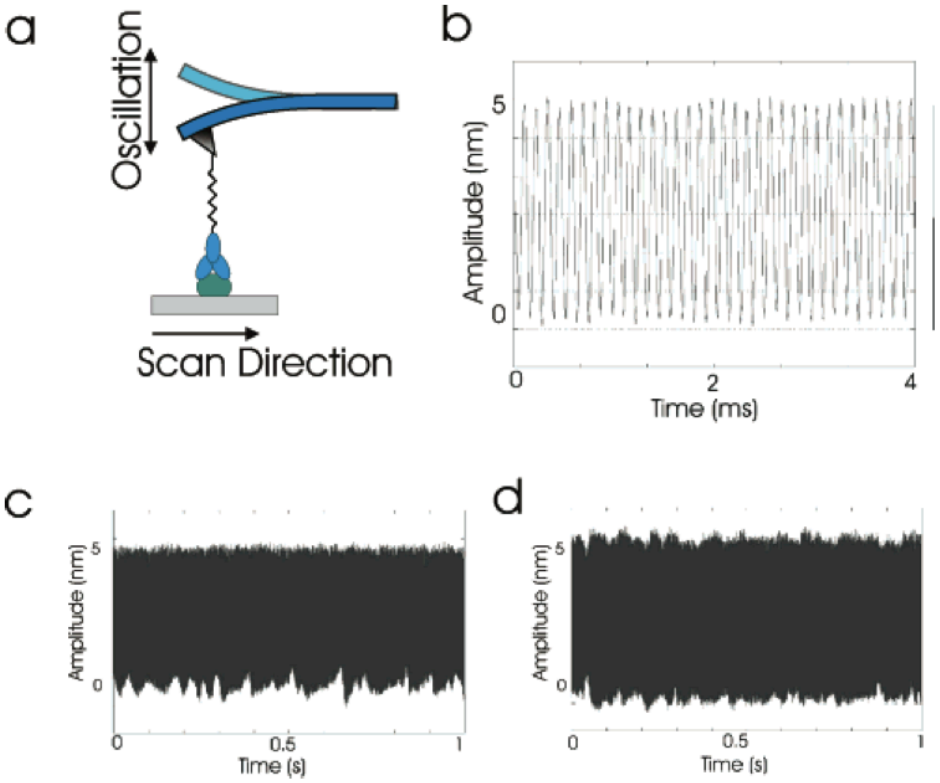


Figure 11 . Principles of Recognition Imaging. [67] **a. Functionalized tip. b. Deflection signal of the oscillating tip in terms of amplitude. C. Deflection signal of a bare tip, showing no change in the upper maxima (recognition), but showing changes in the minima giving topographic information. d. Deflection signal for functionalized tip scanned over the molecule of interest, showing changes in both upper maxima and minima of the deflection signal.**

With the functionalized tip, when the antibody tethered to the tip attaches to its cognate on the surface, the upward swing of the oscillation amplitude is hindered due to the stretch in the PEG tether. With the decrease in the peak of the amplitude, the servo comes into action and pulls

back the tip away from the surface and the original amplitude is restored, but the peak is shifted downward to an extent of the previously reduced amplitude. [19] This can be seen in the Figure 11. The lowering of the amplitude peak can be seen in the compressed deflection signal. Thus , maxima and the minima of the z amplitude, gives valuable information with the fluctuations in maxima giving the recognition signal ,i.e., the reduction in amplitude signifies the binding process being taken place ,and the fluctuation in the minima giving the topographic signal. Thus, the simultaneous mapping and the detection of the recognition sites is possible with nm positional accuracy. [20], [61], [67], [68]

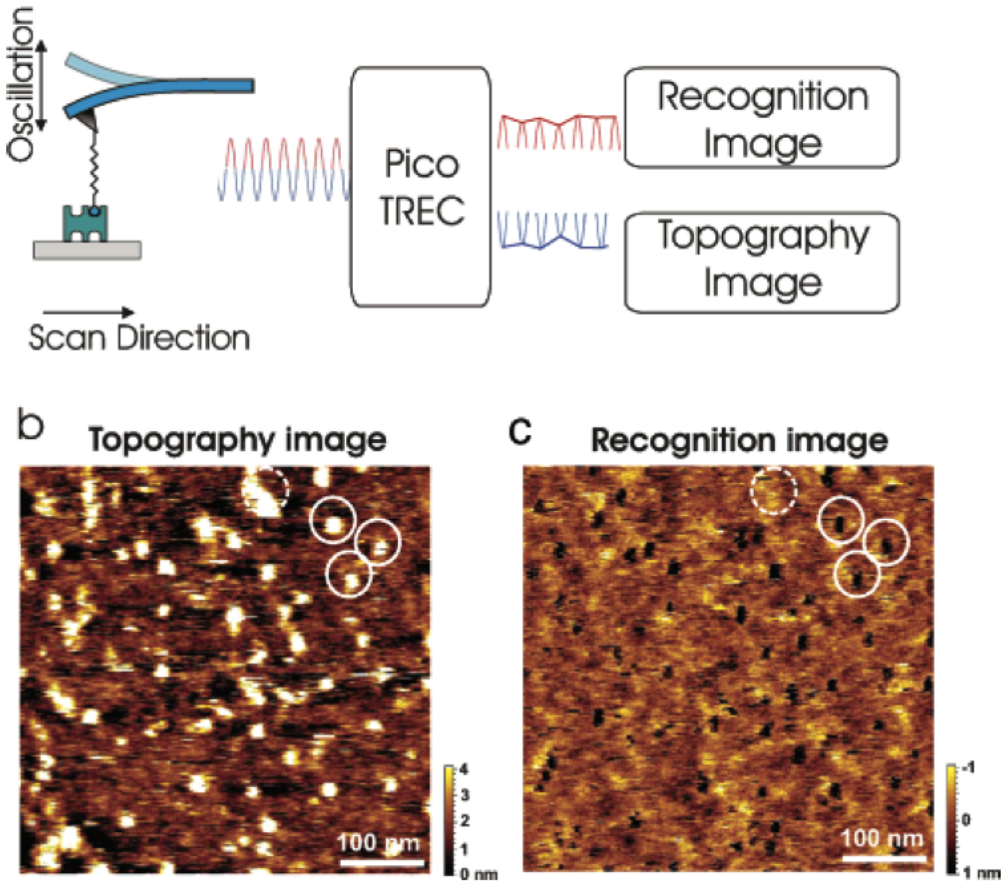


Figure 12. Simultaneously occupied topography and recognition signals [67]. (a) Signal processing for simultaneously obtaining topography and recognition images. The raw cantilever deflection signal of the oscillating cantilever is thereby split via the TREC-box into upper parts (corresponding to the recognition image) and lower parts (corresponding to the

topography image). (b) Topographical image of avidin molecules adsorbed to mica acquired with a biotin-tethered tip. (c) Simultaneously acquired recognition image. A good correlation between topography and recognition was found (solid circles). Topographical spots without recognition denote structures lacking specific interaction (dashed circle)

In the electronics, the maxima and the minima are filtered and amplified by a peak detector. The TREC box separates the upper maxima and the minima from the signal (Refer Figure 12 and Figure 13). U_{down} drives the feedback loop for the topographic image of the sample and U_{up} gives the recognition image. [69]

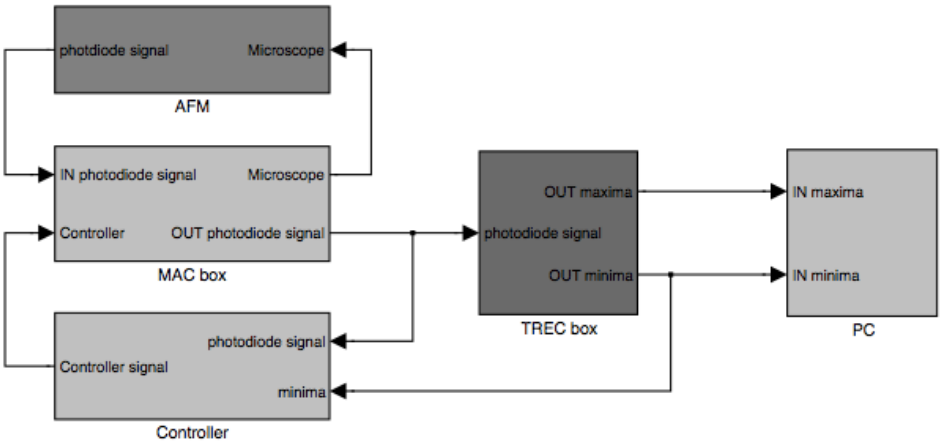


Figure 13. Signal processing for recognition imaging[69]

1.3.4 FORCE DISTANCE CURVE

Interaction forces between single molecule ligand and receptor pairs are measured by force distance cycles using functionalized tips with the ligand on the tip and the receptor on the APTES modified mica surface or any covalent surface.

At a fixed lateral position, the tip approaches the surface and is later retraced. During this force distance cycle, photo detector measures the deflection of the cantilever, which gives the force according to the Hooke's law F=-k.x, where k is the spring constant of the AFM tip.

For a bare tip, when the tip is approaching the surface, i.e. region 1-2, there is no deflection. As it reaches the surface and there is tip-surface contact, the tip experiences repulsive

Vanderwaals forces, leading it to deflect upward. (region 2-3). This repulsive force increases linearly with negative distance and reaches a maximum at the end of 3. Subsequent tip surface retraction from the surface leads to a relaxation in the repulsive force being exerted on the tip, until it reaches a zero force. This can be seen between 4-5. Upon further retraction , tip experiences an attractive force due to the surface forces and then on further more retraction, jumps off the surface (region 5-6).

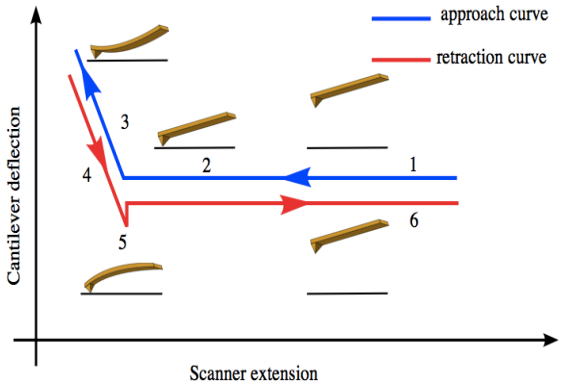


Figure 14. Force distance curve

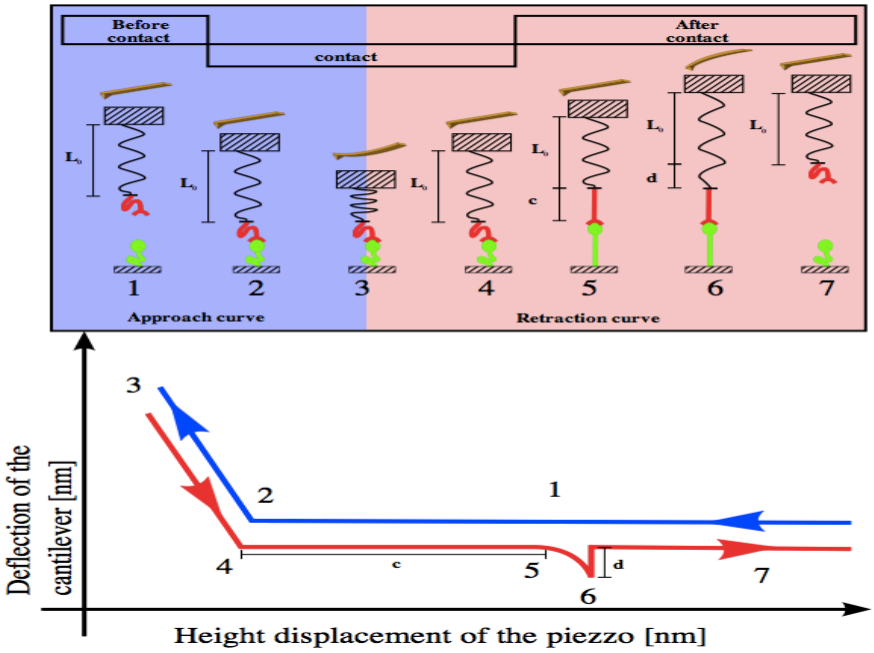


Figure 15. Force Distance curve for a functionalized tip.

In the case of functionalized tips, in the region 1-2, at 2 the functionalized tip comes into contact with its cognate on the surface. It follows the same steps as the original force curve and approaches the surface, experiencing vander waals forces of repulsion and reaches a maximum with negative distance .Subsequently the tip gets retracted from the surface, leading to relaxation in the repulsive force being exerted , until it reaches zero force(region 3-4) . On further retraction from the surface, tip experiences an attractive force due to the attachment of the ligand and the receptor as seen in region 5-6. This attractive force keeps on increasing as the tip –surface distance keeps on increasing and the tip keeps on bending downward more and more. This shape of the curve, which is nonlinear, is governed by the elastic properties of the PEG tether. This represents the stretching of the PEG tether and is parabolic and reflects the increase in the spring constant of the cantilever during the stretching of the PEG tether. This specific shape of the stretch of the PEG tether corresponds to a single molecule recognition event. The strength of the interaction is governed by the binding between the ligand and the receptor. When this interaction has reached the maximum possible force, the receptor –ligand complex dissociates and tip jumps back to its resting position 7. This gives a measure of the unbinding force between receptor and its cognitive ligand. If there is no binding between the receptor and ligand, the trace and the retrace looks the same.

With all these concepts and basic functionalization, there have been many recent progresses and applications of AFM done in the field of biotechnology. Let's take a review of it.

1.4 RECENT ADVANCES IN AFM RECOGNITION IMAGING

1.4.1. AFM IN LIVE CELL IMAGING

The most recent, challenging and helpful application of AFM in world of medicine is the mapping of receptors on the cell surfaces. This can give knowledge in vivo. Many other methods including immunostaining have been utilized in this field but most of them lack in the level of resolution they can give. AFM gives nanoscale resolution of both topographic and recognition data and also can give an idea of the interactions present between the molecule of interest and the cells. .

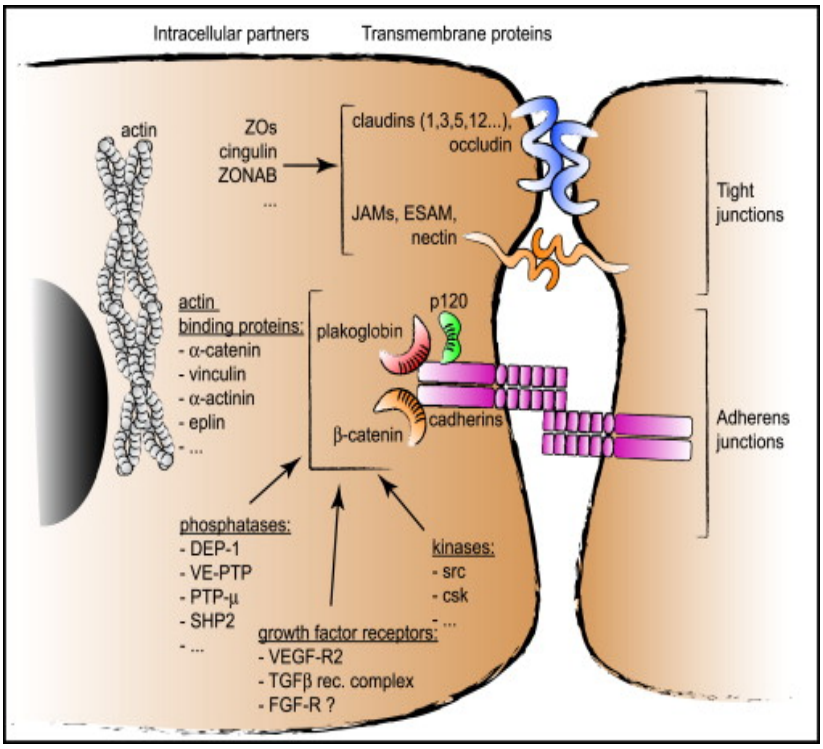


Figure 16. Cadherins in cell-cell junctions.[67]

One of the most recent works done on this include the study done on the Vascular Endothelial (VE) cells.[50] Some of the human vascular pathologies like ederma, vascular malformations and hemorrhagic stroke are associated with defects in the endothelial permeability. In most of these cases, the cell to cell junctions are disrupted leading to the detachment of the endothelial cells from the vessel walls. A study related to the intracellular junctions can play a vital role in understanding the basis of these pathologies. In the process of transferring intracellular signals through the junctions , a major role is played by the transmembrane adhesion protein, which is cadherin in majority of the endothelial cells.[67] VE cadherins are very cell specific, making it very clear that they are required for selective cell- cell recognition. The formation of clusters of cadherin for the organization of the cell junctions is a very crucial process for the cell-cell signaling .This cluster formation further leads to the interpretation that this is required for the strong binding between cells. The drastic increase in the permeability and reduced VE-cadherin mediated cell-cell binding on depolymerization of the actin filaments supports the fact that these proteins are linked through their cytoplasmic domain

to the actin filaments.[68],[69] Also, it has been observed that the VE – cadherin transinteraction is very low and thus the cadherins have to tether to the cytoskeleton(actin) ,so that they rebind after dissociation. The basic mechanism underlying this linkage was not very clear though. With this AFM study, there was an effort to map the receptors for an understanding of the dynamics of receptor localization with the cell membrane topographical features. (actin filaments).

Molecular recognition studies using VE cadherin –Fc molecules as a sensor on the tip were done on the MyEnd (Mouse myocardium) cells. MAC tips were functionalized with the VE –cadherin Fc molecules using NHS-PEG- aldehyde linker (8nm). The cells were gently fixed with 0.5% Glutaraldehyde and incubated at 37^0C overnight. All the experiments were conducted in Hank's balanced salt solution (HBSS, 1.8mM Ca^{2+}).

Topographic data of the MyEnd cells showed distinct highly elongated features, with linear and branched filaments, and they had whirl like organizations. The images were comparable to the earlier ones obtained through light microscope.[70] The fixation procedure for the cells used here conserved the filamentous structure without dehydrating them, which is normally a very big drawback of fixing the cells. This greatly helped in imaging them as they became much stiffer. It also helped in stopping the lateral diffusion of the receptors, which helped in doing recognition imaging on them. The treatment with nocodazole to depolymerise the microtubules , preserved the filamentous structure of the cell cortex.

As VE-cadherins are supposed to be located along the edges of the cell-cell junctions, the recognition maps were taken around the same area in HBSS calcium rich medium. The TREC images showed microdomains with the dimension range of 10- 250 nm.

The size distribution (Figure17) gives a clear indication of two populations of the microdomains: larger domains in the range of ~80 nm and ~180 nm and smaller ones, with a very small frequency,in the range of ~ 10-20 nm, clearly implying the clustering of VE cadherins at the

cell-celljunctions.

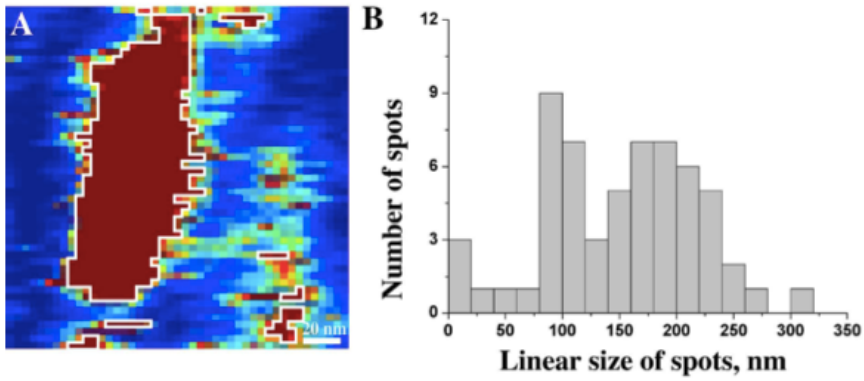


Figure 17. Size distribution of the recognition maps.

Specificity experiments were done in the absence of Ca^{++}, and showed the abolished recognition signal supporting Baumgartner observation that the cis dimers on the tip dissociate into inactive monomers, making it inactive. The simultaneous topographic and recognition data figure (18) showed the presence of actin filaments near the recognition spots, i.e., cadherin, supporting the hypothesis of VE cadherin attachment to the f-actin cytoskeleton.

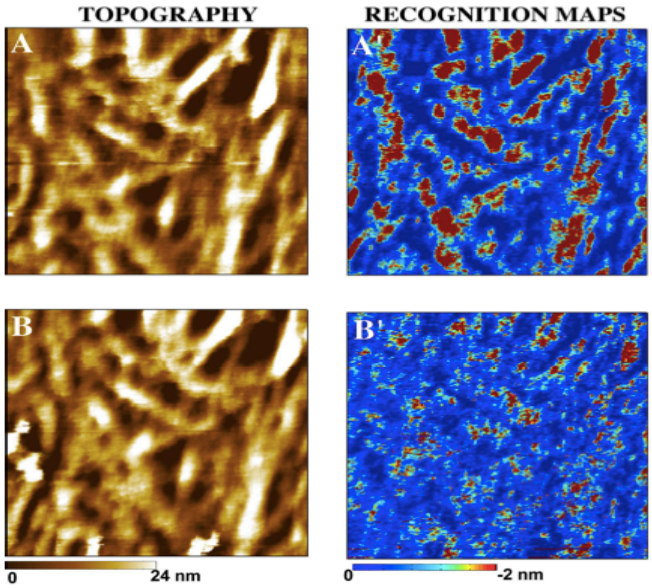


Figure 18. Simultaneous topography and Recognition images obtained from MyEnd cells (fixed with GD) [50]

This work clearly demonstrates the advantage of TREC over other techniques. No technique in the biological world is able to give the simultaneous topographic and recognition data at such a high resolution of 5nm.It is successfully able to demonstrate the anchorage of actin cytoskeleton to the cadherin molecules, also giving an insight into the location and clustering of the cadherin molecules at the cell-cell junctions.

Van Vliet and colleagues conducted a study in 2007[71], on HUVEC (Human umbilical vein endothelial cells) cells on both living and fixed cells. These cells endogenously express VEGFR2. MAC mode tips were functionalized with Mouse antihuman monoclonal anti-vascular endothelial growth factor receptor 2, tethered with PDP-PEG-NHS ester. The recognition images of HUVEC cells taken by these functionalized tips, overlapped on the topographical data, gave critical information on the distribution of the receptors in the cells and their correlation with cytoskeleton and other structural features in the cell. The diameter of the recognition spots was measured in the range of 45.9 ± 8.9nm. Specificity was demonstrated through competitive binding

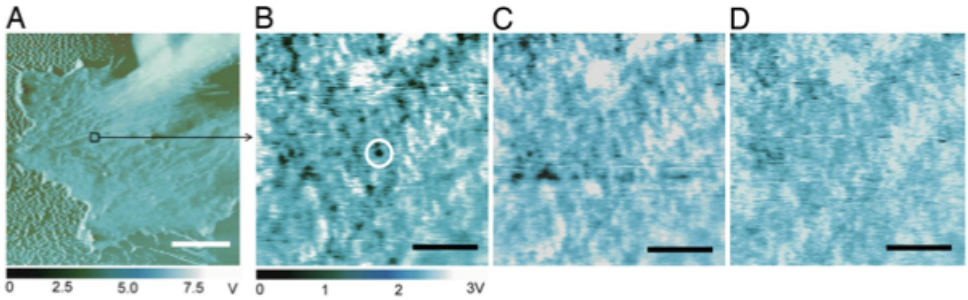


Figure 19. Chemomechanical imaging of individual VEGFR2 receptors on fixed HUVEC surfaces. (A) Phase image of cell body and periphery. (B) Recognition image over indicated area in A shows strong binding events between the anti-VEGFR2 functionalized probe and the cell surface as discrete, dark spots (e.g., circled) that are ostensibly VEGFR2. (C and D) The capacity to block these binding events by addition of 5 g/ml soluble anti-VEGFR2 supports this binding specificity between the probe and VEGFR2, as the number of recognition events decreases with time postblocking of 12 (C) and 60 (D) min.

as seen from figure 19. The recognition image shows an average of 1.47± 0.38 x 10^5 VEGFR2 per cell. The number was in good agreement with the measurement done using radiolabeled ligands, 1.1 x $10^{5[72]}$ and 1.5 x$10^{5[73]}$ respectively, on HUVEC cells. The receptors seem to be distributed in a very nonuniform fashion. The images also show a close cytoskeletal association of these receptors. The distribution was done wrt the centre and apparent edge of subsurface cytoskeletal bundles.61 % of the recognition spots were observed right adjacent to the underlying cytoskeletal bundles; 34% were located within a range of 72± 49nm from the bundle edges; and lastly almost 5% were at more than 500nm from the bundle edges. This observation supports the view that VEGFR2 is very closely related to transmembrane integrin complexes that transmit the response from the extracellular matrix to the actin cytoskeleton.

An average over force displacement retraction profiles between the probe bound antibody and cell surface receptors gave ligand binding affinities on individual cells. The distribution of these forces gave two maxima at 32.5±2.5 and 64.1±5.4pN, indicating presence of binding between two receptors with a single antibody or binding of two antibodies on the probe to closely spaced receptors and gave an off rate of k_{off} = 1.05±0.6 x$10^{-1}s^{-1}$.

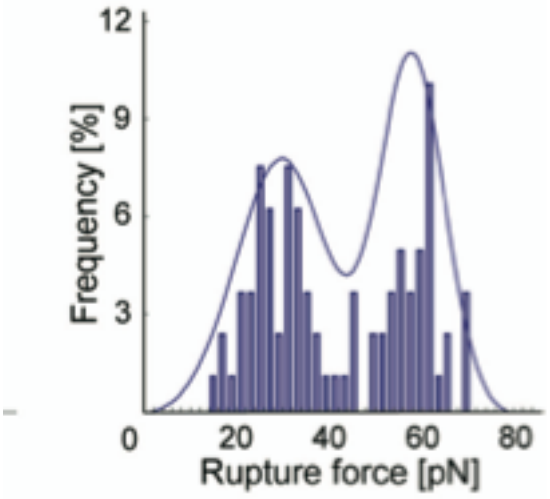


Figure 20 Distribution of rupture events showing two maxima.

Thus, in this study, it clearly shows that these VEGFR2 receptors are spatially associated with the underlying cytoskeleton. Also , further studies based on these techniques will not only

enable predictions of the not so well know dynamics interactions between extracellular molecules and the receptors moving on the living cell surface , but also give valuable information on the ligand based clustering and the associations of the cytoskeleton with these receptors . The kinetic constant information extracted from these experiments give an estimate of the speed, specificity and strength of the interactions between the drug to be tested and the receptors.

1.4.2 AFM IN THE STUDY OF TRANSPORT MECHANISM THROUGH CELLS

Jens Christian Skou, [74]discovered the Na^+-K^+ pump in 1950 and was awarded the Nobel Prize for the same in 1997. The understanding of the phenomena marked a milestone in the understanding of membrane transport. It gave an insight on how ions get into and out of the cells. The Na^+-K^+ ATPase helps maintain membrane resting potential by pumping out 3 Na^+ and moving in $2K^+$ ion, thereby maintaining a positive potential wrt to the inside of the cell at the expense of ATP. It also helps in the transport mechanism through the cell. Export of Na^+ using this pump against the gradient at the expense of ATP, serves as an active transporter to import glucose, various amino acids and other nutrients. It is further required for the maintenance of cell volume and for signal transduction. [75]

Abnormalities in the function or the number of this pump can lead to several health problems, particularly heart diseases and hypertension. Several of the heart failures are a result of the significant reduction in myocardial concentration of Na^+-K^+ ATPase. Excessive renal absorption of sodium has also been associated with hypertension. Thus, exploring the location of Na^+-K^+ in cell membranes as well the relationships it has with different membrane proteins can help in the treatment related to Na^+-K^+ abnormalities. Hongda[76] and colleagues explored the Na^+-K^+ pump and were able to localize ATPases in quasi-native cell membranes(RBCs) using TREC imaging.

Human RBC's were sheared open after attaching them to the APTES mica surface using low salt PBS buffer.[77] Presence of Na^+-K^+ ATPases in the inner cell membrane was verified by attaching QD's functionalized anti-ATPase . Figure (21) shows the presence of the abundant Na^+-K^+ ATPase. For TREC images, MAC tips were functionalized with anti-ATPase with a

heterobifunctional PEG linker. The recognition images showed a uniform distribution of Na^+-K^+ ATPase with only 10% aggregation. Thus, supporting the view that these are present all over the membrane and not localized.

The superimposed TREC data revealed that the average height of the proteins with the recognition signal was in the range of 12-14nm.

The interaction between Na^+-K^+ ATPases and anti-ATPases was further measured by the force curve measurements. The distribution shows a peak of 80pN at a loading rate of 49nN/s. This agrees well with the typical antibody-ligand data range. The specificity for experiment was established by imaging the same sample area again and by adding ATPase in the flow cell while taking the recognition signal. The recognition signal was lost, thus confirming that the interaction is specific.

1.4.3 AFM IN THE STUDY OF CHROMATIN EXTRACTED FROM CELL LINES

Since the discovery of the structure of DNA in 1953 by Watson and Crick [78], DNA has been the topic of paramount interest for all the scientists in the world. Now, almost 50years later, it is realized that the development and phenotype expressed by any cell mainly depends on the information encoded in the DNA. Though, the code inherited lies entirely in the DNA, but its expression depends on the modifications of the DNA and how the DNA is wrapped or packaged into chromatin. Some of the genes get silenced and some of them are expressed. The patterns of gene expression, thus greatly depends on how the DNA is packaged into chromatin. Thus, inorder to study gene expression, study of chromatin is very important. And, if the direct visualization of chromatin can be done along with mapping its interactions, it can give us deeper insights involved in the process of gene expression. With AFM, it is possible to map and localize the nucleosomes at nm resolution.

147 base pairs of DNA are wrapped in 1.67 left handed superhelical turns around the histone octamer, which comprises primarily of 2 copies each of the core histones H2A, H2B, H3 and H4.The study of these histone proteins and their modifications and variants can be very crucial in understanding gene expression. [79]

Hongda and colleagues [41] studied one of these modifications using AFM. CenH3 is a centromere specific H3 histone variant, which is required for the spindle attachment during the process of mitosis and meiosis. Previously, it was believed that the centromeric nucleosomes were the same as the normal histone octamer nucleosomes. But, the studies conducted in 2007[41], very

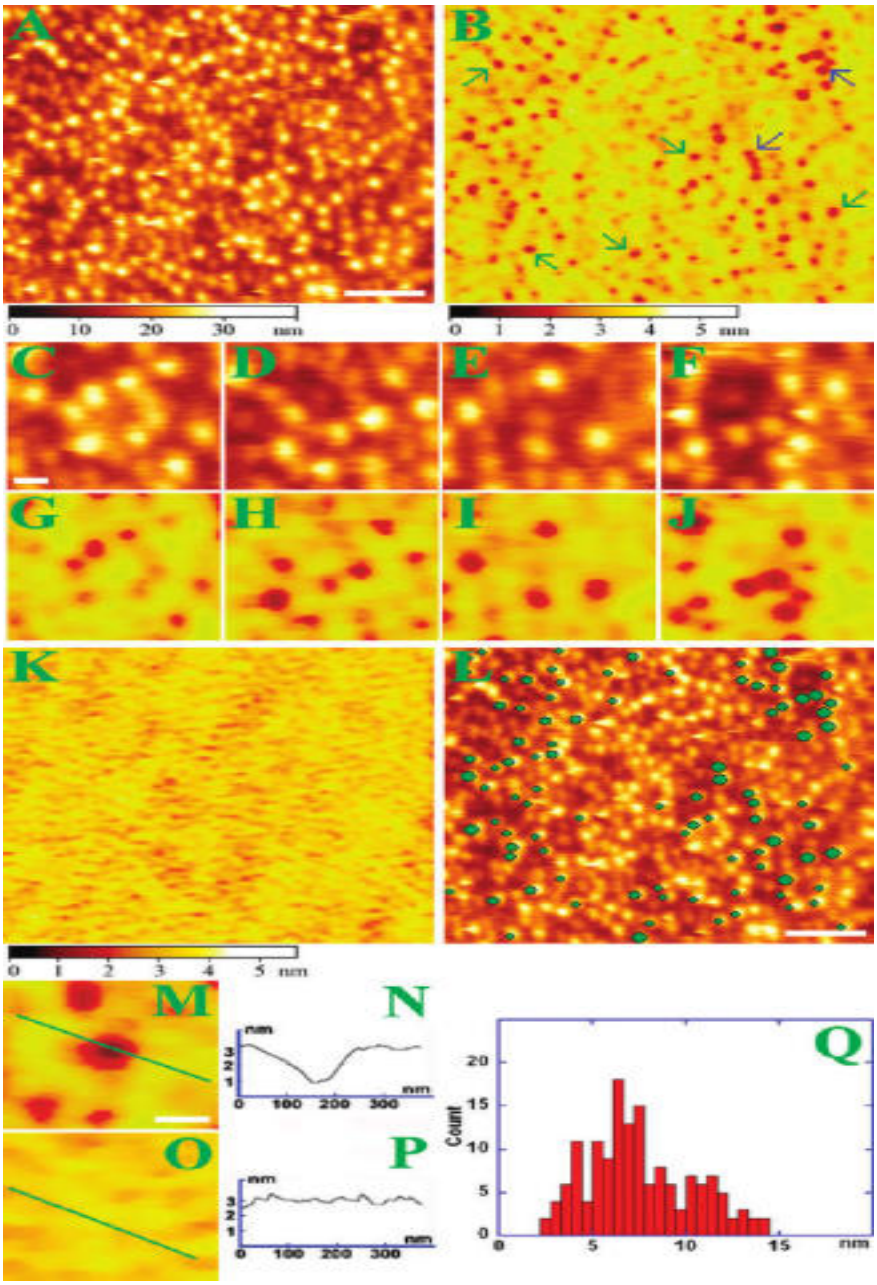


Figure 21. (A) Topographic image of the inner leaflet of cell membranes. Scale bar is 500 nm. (B) Corresponding recognition image of Na+-K+ ATPases (dark spots). Red and blue

arrows indicate isolated Na+-K+ATPases and Na+-K+ATPases aggregations, representively. (C-F) Magnified topographic image of the inner membrane. Scale bar is 100 nm. (G-J) Magnified recognition image corresponding to figure 3C-F. (K) Recognition image of Na+-K+ ATPases after blocking by free anti-ATPases. (L) Topographic image of the inner cell membrane with recognition spots superimposed by overlaying them (green dots) on the top of the topographic images. Scale bar is 500 nm. (M) Magnifying recognition image from Figure 3B. Scale bar is 100 nm. (N) Cross section analysis along the line in Figure 3M. (O) Magnifying recognition image from Figure 3K. (P) Cross section analysis along the line in Figure 3O. (Q) The height distribution of proteins in the inner leaflet of cell membranes.

evidently predicted that the cenH3 nucleosomes were about half the height of the bulk nucleosomes, with one molecule each of the cenH3, H4, H2A and H2B. Figure (22). To make sure that the results were not for an immature intermediate of cenH3 and were rather matured cenH3, another study was conducted in the same year [42]. CenH3 enriched nucleosomes were Glutaraldehyde fixed and immobilized on the hydroxylaptite. This is a DNA binding step, to be sure that only stable nucleosomes are used in the experiment. The nucleosomes were then washed with 0.35M salt to remove the non-DNA nucleosomes and then eluted out (only histones) using 2M salt. The height measurements between cenH3 histones and the canonical histones were consistent with the previously evaluated data. [41]

TREC data was collected with MAC tip functionalized with the anti-CenH3 antibody. The simultaneously occupied topography and recognition data is in excellent agreement with each other. [Refer Figure (23)]. 92% of the total cenH3 molecules were identified successfully using recognition imaging. Blocking of the tip with cenH3 peptide, completely removed the recognition signal, thus implying that the recognition process is very specific. Specificity was also established the TREC data with anti-CenH3 antibody over the canonical histone octamer. by taking Even though, they were twice the height of the cenH3 histone, but still there was no recognition signal, clearly distinguishing cenH3 with the normal histone octamers. The dissociation force was also calculated between the cenH3 octamer and the anti-cenH3 histone and the force (63.1 ± 20.6pN)

values were in the range of the normal antibody ligand values.

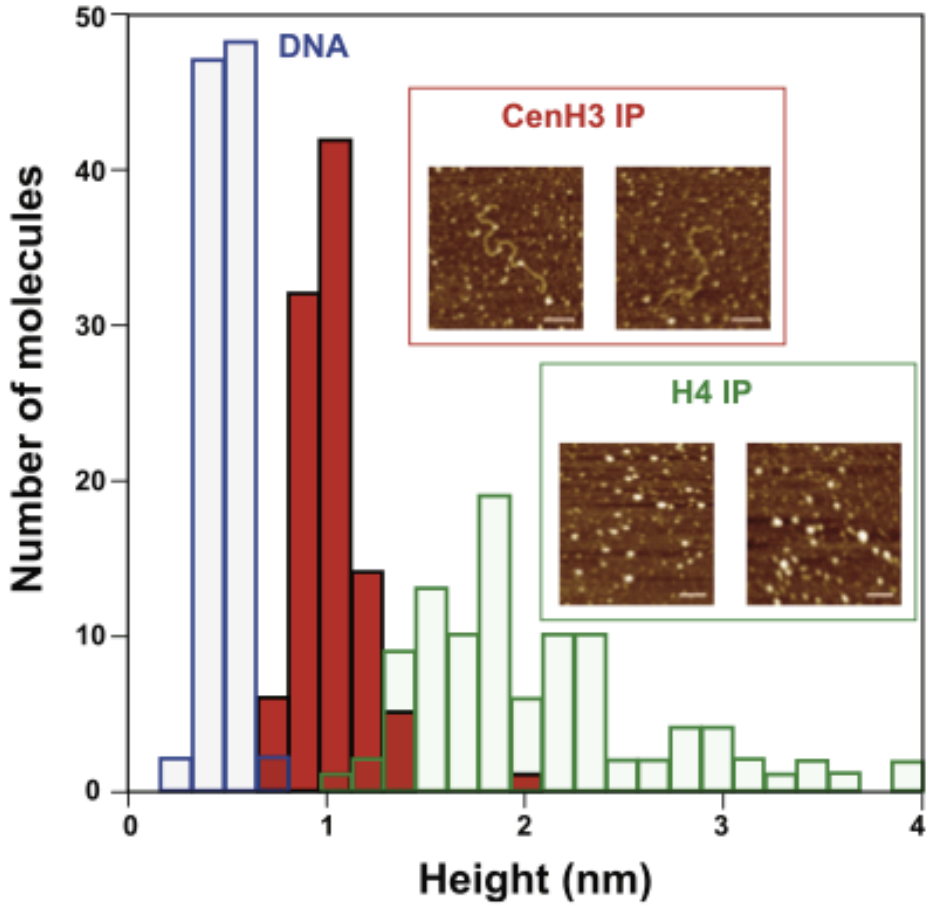


Figure 22 CenH3 Nucleosomes Are Half the Height of Bulk Nucleosomes as Measured by AFMHistogram depicting mean heights of control H4 IP nucleosomes (green bars) and CenH3 IP nucleosomes (red bars) computed from 100 counts of each. Mean H4 IP nucleosomal height is 2.05 6 0.62 nm, in accordance with previously observed heights for bulk octameric nucleosomes. Mean CenH3 IP nucleosomal height is 1.03 6 0.17 nm, or one-half octameric nucleosomal height, consistent with a tetrameric structure. Insets depict representative CenH3 IP molecules (inset, red box) and H4 IP molecules (inset, green box) used in this analysis. DNA heights (blue bars) were computed as an internal control for CenH3 IP nucleosomes. Scale bar represents 100 nm.

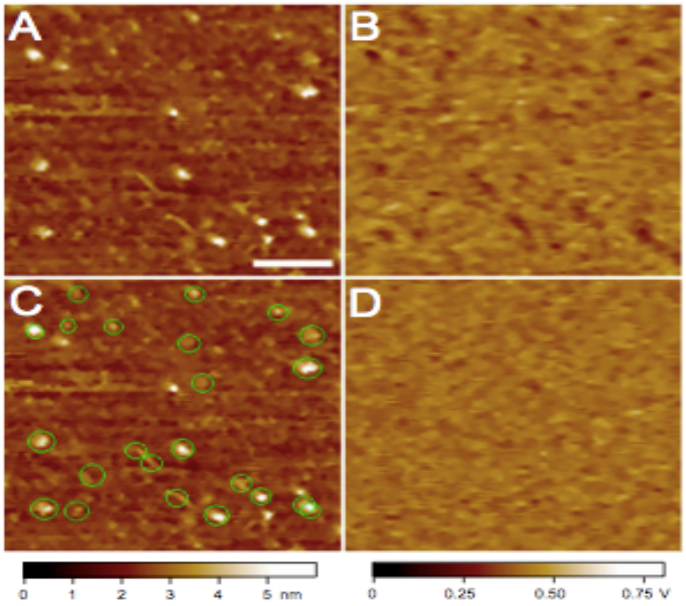


Figure 23. Recognition imaging of native CenH3 core particle complexes. CenH3 core particles obtained by hydroxylapatite enrichment of chromatin followed by immunoprecipitation were visualized by an anti-CenH3 atomic force microscopy tip (A). The recognition signal is detected as dark spots in the recognition image (B). In the overlay shown in (C), the recognition signal is marked with green dots for visual clarity using custom software. No recognition signal is seen after incubation with a peptide corresponding to the CenH3 epitope (D). The bar is 100 nm and the *Z* range and volt scale are shown at the bottom of the images

Thus, this experiment was able to give deeper insights into the histone H3 variant, cenH3. present at the end of the 125 aa N-terminal tail, thereby helping in elucidating epigenetics and their role in gene expression.

1.4.4 AFM STUDY IN ANTIBODY UNFOLDING AT DIFFERENT LOADING RATES

As seen from all the applications of AFM molecular recognition techniques, especially force spectroscopy [80] and recognition imaging [81], an antibody is pulled away from its target protein at timescales very less as compared to their normal interaction time. In the study conducted

by kaur et al, on the dynamics of interactions between antibody and antigen during force spectroscopy and recognition imaging experiments, it was seen that during these pulling experiments in recognition imaging, which is always at high loading rate (confirmed by the onrate data, refer Figure (26)), the antibody frequently unfolds and folds, indicating that the antibody temporarily transforms to a metastable state.

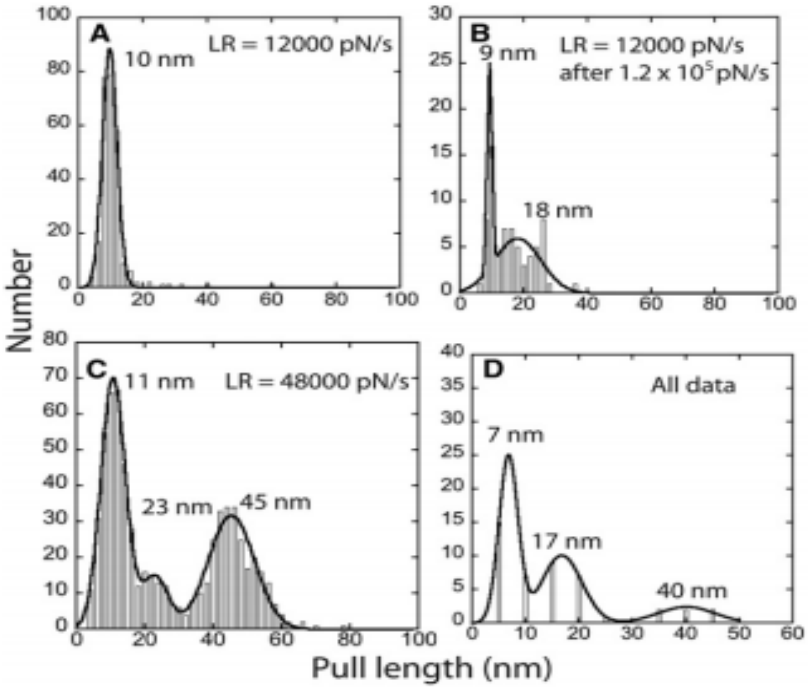


Figure 24. Distribution of pulling lengths showing how high loading rates drive unfold

This was confirmed by the force spectroscopy measurements where an additional force peak was seen at high loading rates, indicating the folding of the antibody, but no such peak was seen at low loading rates. (refer Figure 24). In the recognition image, due to this weak interaction between the unfolded antibody and its cognate as compared to its normal interaction, the unbinding takes an order of millisecond timescale, [80], leading to rapid switching noise in the recognition images. This switching noise in the recognition spots can be used to filter the data out for the weak interactions, arising as a result of higher pulling velocities, leading to the breakdown of the structure. As the interaction between the metastable state and the protein is very weak, this calls for strict specificity controls for the process of recognition imaging. The current specificity

controls although are good, better and reliable results can be achieved with stable rigid recognition reagents, like cyclotides. [83]

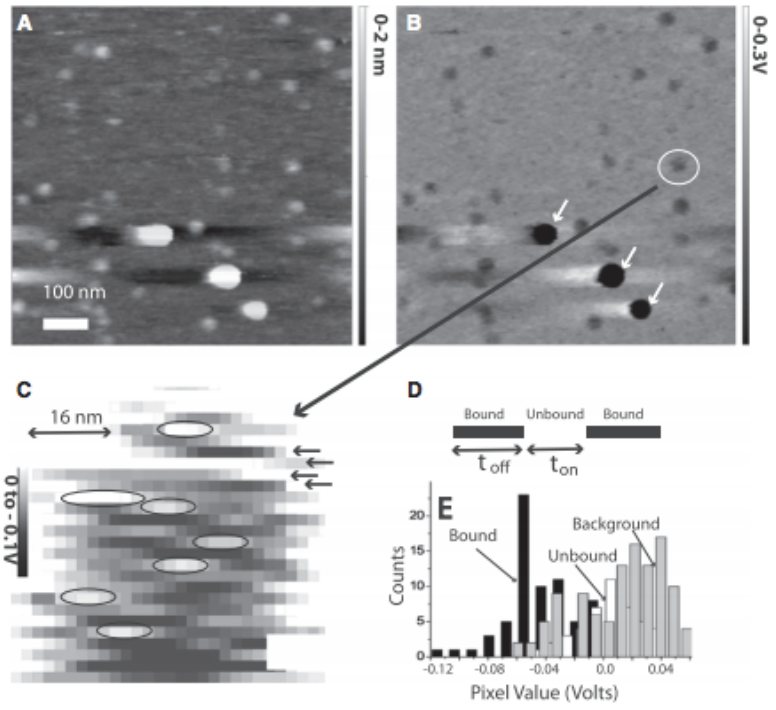


Figure 25. Unbinding events in recognition imaging. A. Topographic Image of receptors B. Simultaneously Acquired Recognition C. Blow up of Recognition Image. D. Interpretation of the fluctuations in the blow up image E. Histogram of Voltage levels in the recognition events.

TABLE 1 On- and off-rates

Sample	K_{on} (1/Ms)	K_{off} (1/s)	K_D (M)	χ^2 (RU^2)
Kinetics of antibody binding using SPR				
Whole IgG	$7.57 \times 10^{+5}$	4.19×10^{-5}	5.53×10^{-11}	0.219
	$4.21 \times 10^{+5}$	2.88×10^{-5}	6.85×10^{-11}	0.715
Fab	$1.39 \times 10^{+6}$	6.93×10^{-3}	5.00×10^{-9}	0.36
	$4.71 \times 10^{+5}$	3.02×10^{-3}	6.40×10^{-9}	0.30
Kinetics of antibody binding from recognition image noise				
Whole IgG	$8 \pm 1 \times 10^{3}$	$2.1 \pm 0.3 \times 10^{2}$	0.1 ± 0.1	

Rates were determined by Biacore SPR (GE Healthcare) for the whole IgG and the Fab fragment and for the whole IgG by recognition imaging switching noise.

Figure 26. ON and off rates calculated for Biacore SPR and Recognition imaging.

CHAPTER 2

SURFACTANT ACTION OF THE METHYLATED DNA IN SILENCING GENE EXPRESSION

PROJECT 1.

2.1 INTRODUCTION

The basis of life of a living organism depends on the ability of its cells to store, maintain and interpret the data in its DNA (deoxyribonucleic acid), the DNA which is contained within the tiny space of a cell. In the early 1940's, because of simple chemistry of the DNA, it was difficult for the scientists to accept DNA as THE genetic material. But, with the experiments conducted by Griffith in 1928[84] and followed by Avery in 1944[85], it was firmly established that DNA was the missing link in carrying and transfer of genetic and hereditary information. The discovery of the structure of DNA in 1953, by Watson, Crick and Franklin[86],[87], further paved the way for the understanding of how the genetic information can be copied and replicated and how it can encode for the instructions for making proteins.

With the exception of retro viruses (genetic material as RNA), all the organisms use DNA as the carrier of hereditary information. The DNA segments or their stretches, which carry this information and thus encode for a protein or for RNA, required for the functioning of the organism, are termed as genes. The term gene was coined from the word "*pangen"* by the Danish botanist Wilhelm Johannsen in the year 1909.

A DNA molecule consists of two strands of polynucleotide chains folded into a helix. It consists of a sugar-phosphate backbone (bonded covalently), with the nitrogen bases of the two separate strands hydrogen bonded with each other. These hydrogen bonds between the two bases hold the two strands together. Nucleotide consists of a five-carbon sugar, attached to phosphate groups on both the ends with a nitrogen containing base. In the case of DNA, the sugar is deoxyribose, while in the case of RNA, it is ribose. The bases may be either adenine (A), guanine (G), thymine (T) or cytosine(C). The sugar and phosphate, covalently attached to each other via phosphodiester bond, forming the backbone of the DNA. [86]

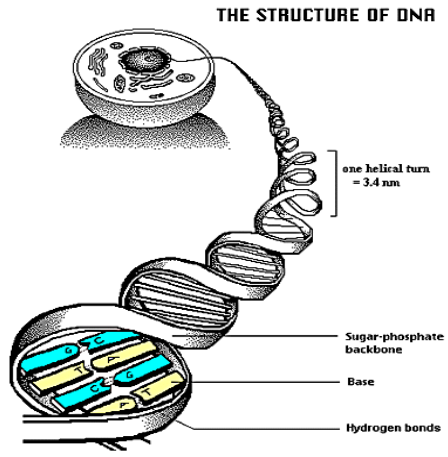


Figure 27. Structure of DNA [88]

Eukaryotic organisms store genetic information in their DNA inside the nucleus of the cell and some of it in organelles like mitochondria or chloroplasts, whereas, the prokaryotes store their DNA in the cytoplasm in the form of circular DNA, called plasmids. But, the basic structure of DNA remains the same for both. The 3D structure of the DNA as predicted by Watson and Crick is based on the chemical and structural features of the polypeptide chains coiled into helix.

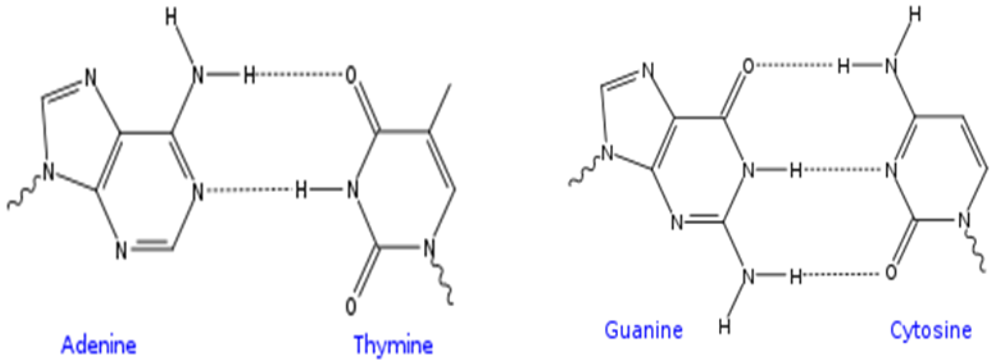


Figure 28. Nitrogen Bases in DNA

The two polypeptide chains are held together by hydrogen bonds between the nitrogen bases on the two different strands and thus, the bases always are situated towards the inside of the helix with sugar- phosphate backbone on the outside. The bases are the nitrogen containing ring compounds, the bulkier one with a double ring named purines and the single ring ones as pyrimidines. Adenine, guanine are the purines and cytosine, thymine and uracil are pyrimidines. A

always pairs with T, and G with C, i.e., generally purines form hydrogen bonds with pyrimidines. These base pairs form different number of hydrogen bonds with A = T and G ≡ C. Thus, in this arrangement, all the base pairs are of similar width and hold the sugar phosphate backbone at equal distance from all of them. The two strands run antiparallel to each other, coiled around the same axis, in the helix and thus each strand contains complementary bases to its partner strand. It forms a right handed helix; with one complete turn every ten base pairs, centre to centre distance of 3.4 nm and a radius of 1nm. [86], [87] The backbone of the two strands of DNA is close together on one side as compared to the other. The region of the backbone, where the two strands are far apart forms the major groove and the region of the backbone where they are the closest forms the minor groove. As the backbones are farther apart in the major groove, this is the site for the action of DNA binding proteins and modifications (on the nitrogen bases which are towards the inside) as it is easier to be accessed.

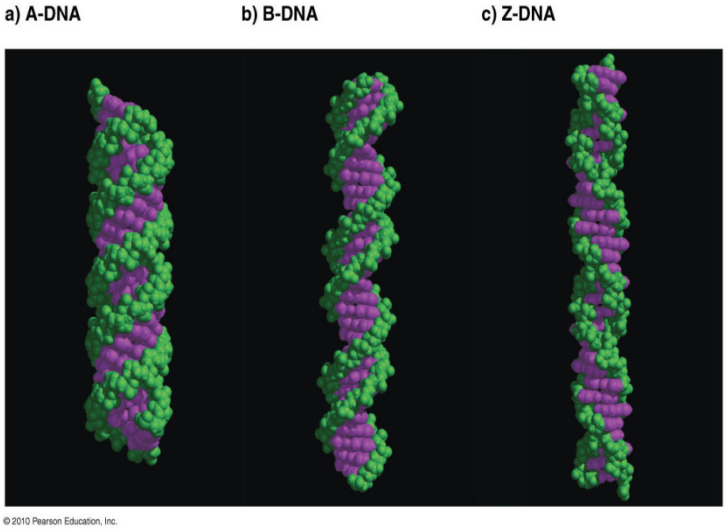


Figure 29. Different forms of DNA [89]

DNA described by Watson and Crick was the B form of DNA. It is now known that there are other forms of DNA (Refer Figure 29). Table 3 gives a brief description of the different forms of DNA.

	A -DNA	B-DNA	Z-DNA
Helix sense	right	left	Left
Rotation /bp(^{o})	33.6	35.9	30
Repeating unit	1 bp	1 bp	2 bp
bp per turn	11	10.5	12
Diameter	2.3 nm	2.0 nm	1.8 nm
Rise/ turn of helix	2.46 nm	3.32 nm	4.56 nm

Table 3 Different forms of DNA

2.2 FUNCTIONS OF DNA METHYLATION

The modifications of histones and chromatin can lead to the silencing of genes, but these processes are reversible and cannot be utilized for the long term silencing. In order for the prolonged mechanism, one of the important components utilized by mammalian cells is DNA methylation. This covalent modification of cytosine to 5 methylcytosine, is not only very stable, but can be passed on to the generations through cell division.

DNA methylation is also involved in x-chromosome inactivation and gene imprinting. This is the exceptional case where the CpG islands (refer Section 2.4.1) are methylated. In both of these phenomena, one copy of the chromosome becomes methylated leading to the expression of the active one. X-chromosome inactivation occurs during embryogenesis and one of the alleles is silenced and stabilized in the repressive, inactive state. Genomic imprinting is required to determine which parental allele will be expressed. It also plays a major role in the regulation of transcription. Normally, this regulation is done with the attachment of mecp2 (methyl CpG binding protein), to the methylated CpGs. With the recruitment of methyl binding proteins by the m-DNA , there is inactivation of the chromatin due to the binding of mecp2 with the transcription repressor sin3A and HDAC(Histone Deacetylase) leading to the condensed form of the chromatin and thereby, preventing the transcription factors to bind to the DNA [90]. DNA methylation can also affect the nucleosomal occupancy at the transcriptional start sites of the genes which can

further affect the transcriptional activation of the genes. With the nucleosomal occupancy being affected, there is a decrease in the binding of transcription factors. Normally, the promoters contain nucleosomal free regions for the binding of the transcriptional activators.[91] With the recruitment of methyl binding proteins to attach to the m- DNA , these methyl binding proteins further attach to the Brahma, which is the catalytic subunit for the attachment of the chromatin remodeling complex SWI/SNF. [92]. Thus, with the chromatin remodeling complex attached to the methylated sites, there is obviously a tendency for changes to the nucleosomal occupancy.

2.3 HISTORY OF DNA METHYLATION

Although, the phenomena of covalent attachment of the methyl groups to the nucleotides was already described in the late 1940's and early 1950's by Hotchkiss and Wyatt[93], but, it was not until the discovery of restriction endonucleases in bacteria by Daisy Dussoix and Werner Arber in 1962[94] , that the biological importance of these modifications was known. By 1970, with the development of restriction endonucleases by Hamilton Smith and group [95], it was appreciated that these enzymes whose activity was reduced in the presence of 5-methyl cytosine could be of great significance in assessing the methylation status of DNA sequence. The observation by Waalwijk and Flavell [96], that HpaII and MspI, both recognize the cytosine in the DNA sequence, but, cannot do so with 5-methylcytosine, gave a very basic mechanism for the analysis of methylation patterns in DNA. Further, in the early 90's ,with the introduction of bisulfite sequencing technique by Marianne Frommer,[97] there emerged a 'golden standard' for the analysis of DNA methylation. The method, though is very reproducible and precise, but is very expensive and laborious.

2.4 MECHANISM OF DNA METHYLATION

Human genome consists of approximately 3 x 10^9 bases in total[98], out of which, roughly 2.7 x 10^7 cytosine residues, are located in the CpG dinucleotides ,which are mostly methylated genome wide[99]. With the studies conducted on the methylation of the cytosine residues and its biological importance being realized, it is not wrong to change the often quoted "four bases" to five, now including 5-methylcytosine as the fifth base. DNA methylation is an epigenetic modification

which affects chromosome stability and gene expression but does not change the genetic code. It describes the addition of a methyl group at the C-5 carbon of the cytosine ring, creating 5 methyl cytosine. Normally, methylation is targeted for cytosines which are always next to guanine. With the addition of the methyl group on the 5' position of the cytosine ring, there is alteration in the physical appearance of the major groove of the DNA where after the modifications, the methyl binding proteins attach. As these changes are copied through the cell division process, these result in the heritable changes in the chromatin structure.

For the methylation reaction to take place, the three most important components, the methyl donor, the cytosine to be methylated and the catalytic nucleophile, need to be in close proximity to each other. During the reaction, the DNA is in its B conformation and therefore the bases are buried towards the inside. All the hydrogen bonds between the nitrogen bases are intact during the reaction except for the targeted cytosine. There is a drastic distortion of the GC bond of the targeted cytosine and the cytosine is flipped out of the helical structure for the enzyme to gain access to it. [100] With the cytosine flipped out, the group on the active site of the restriction enzyme(CpG methyl transferase) does a nucleophilic attack on the C6 carbon of the cytosine ring giving rise to a covalent intermediate. Here, C5 activates as a methyl acceptor by saturating the double bond. S-adenosine methionine donates the methyl group to the C5 carbon to form S-adenosine hydrocytosine (Adohcy). As Adohcy is both nonplanar and nonaromatic, it cannot stack in the DNA and is eliminated. With the elimination of the H at the C5 carbon and enzyme at the C6, there is a regeneration of the double bond between C5 and C6 leading to the formation of 5 methyl cytosine. [101]

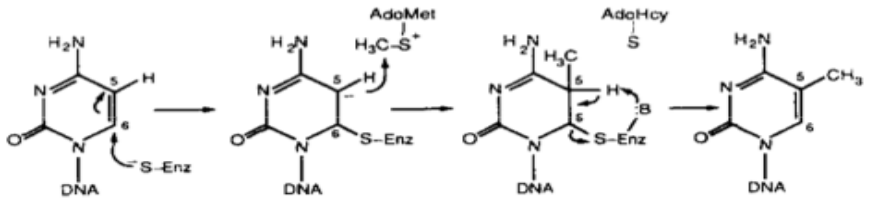


Figure 30. Mechanism of cytosine methylation. [100]

As CpGs are the sites of methylation, let's understand the methylation at these sites.

2.4.1 CPG METHYLATION

"CpG" is the notation used for cytosine- phosphodiester- guanine, i.e., cytosine and guanine covalently bonded by a phosphodiester bond (and not the conventional hydrogen bond), on the same DNA strand. These are sites of methylation in the genome of a living organism. Methylated Cytosines form the fifth base of the DNA with, Adenine (A), Guanine (G), thymine (T) and cytosine(C). But, the point of interest is that why CpGs are so special relative to the other 15 possible base combinations.

According to the density of CpGs, there are two classes of promoters: a. CpG dinucleotides, one where the frequency of CpG is the same as the average of the genome, i.e., 1 every 100 nucleotides b. CpG Islands, where the frequency of the CpGs is 10 times the genomic average. In mammals, dinucleotides are always methylated all across the genome. CpG islands were first called HpaII tiny fragments (HTF), as they contained many sites for the action of the HpaII restriction enzyme. [102], [103] These are normally 300-3000bp long and are CpG rich having G+C content of almost 67%, while the genome averages at about 41%. These are never methylated. [91]

The contrast between the island and the nonisland CpGs is very sharp. The only plausible reason behind this being the fact that the CpGs are not methylated in the islands and thus they always occur at the expected frequency. [102], [91] But, this is not the case for the CpGs outside in the bulk DNA where they are always methylated. Scarano et al. [91] proposed that there is spontaneous deamination of 5-methyl cytosine to form thymine and thus T: G mismatch is generated. As expected, the proof reading for T: G is far less efficient, and thereby leads to mutation and is expressed as TpG or CpA. This leads to the underrepresentation of CpG dinucleotides (21% of the expected frequency). [105]

Chromatin analyzed at these islands revealed a very 'active or open' form. It was due to the hyperacetylation of histones H3 and H4 and absence of histone H1, which was observed in these.

Also, the nucleosome free regions, which are abundantly seen, showed enhanced sensitivity to nucleases, relative to the bulk DNA outside these islands.

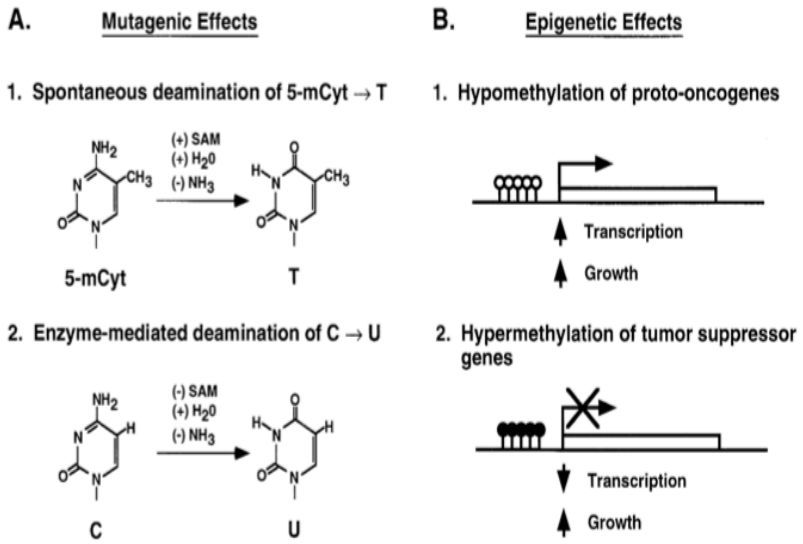


Figure 31 Mutagenic and epigenetic effects of DNA methylation [106]

Thus, chromatin adopts an open structure, and gives direct access to transcription factors to bind to the DNA in CpG rich regions or islands. [107], [108] Mostly, the CpG islands are located towards the extended 5' end of all housekeeping genes and half of the tissue specific genes. For e.g. desmin, hypoxanthine phosphoribosy transferase (HPRT) and retinoblastoma genes. But there are also few located near the 3' end. E.g. Glucose – 6 – phosphate dehydrogenase. [102], [91] It has also been reported that significant portion of the brain or neurally expressed genes are associated with these islands. [91]

The number of these islands in human genome averages around 27,000 as compared to in the mouse genome [104]. About 70% of the human promoters are GC rich. But, the question, which was not clear, was that are promoters located within the CpG islands. Studies were carried on various housekeeping promoters, such as human CDC2 and thymidine kinase to analyze the distribution of CpG sites across them. It was concluded from these studies, that the promoter region is very precisely located between the 5' boundary of the CpG islands and the transcription start site. It was also further confirmed that CpG islands serve as DNA replication origins. This is

also supported by the fact that they have a high density of transcription factors and open chromatin organization. Thus, we can say that CpG islands serve simultaneously as promoters and replication origins.

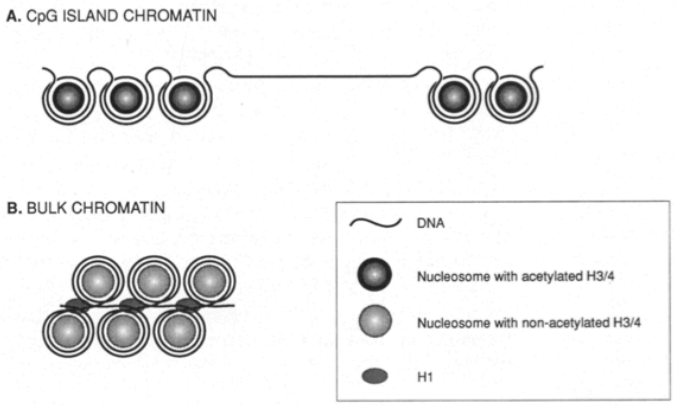


Figure 32. Difference between CpG island chromatin and bulk chromatin..

2.5 DNA FLEXIBILITY

The nucleus forms the heart of the cell and houses the most crucial component of the life, the DNA. The average diameter of the cell nucleus is around 10μ and the DNA, around 2m in length, is packaged very efficiently inside it with the help of the histone proteins. The picture can be seen as packing something into 1/100,000 of its size, implying very clearly that whatever are the phenomena but the flexibility of DNA plays a very critical role in its packaging. The biggest challenge is to understand how the cell controls this process, as the DNA has to be accessed at almost all the stages of the cell cycle and how it deals with the intrinsic semiflexible nature of the polymer, the DNA. The concept that describes this flexible nature of DNA is its persistence length. Persistence length is basically the measure of stiffness of a polymer. It is defined as the length of the DNA, over which the correlations to the tangent, along the direction of the curve are lost. In simple words, the polymers having length shorter than their persistence length behave like a flexible rod and those which are much longer than the persistence length can be described by 3D random walk. The worm like chain developed by Kratky and Porod [109] in 1949, can describe a

molecule with increasing flexibility as its length increases. This model is characterized by the exponentially decaying orientational correlation (C(s)) function which decreases with the increase in the length of the molecule and is given as:

$$C(s) = \exp(-s/l_p) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [2]$$

where l_p is the persistence length of the polymer which describes the decay of the correlation function.

Various methods have been used for measuring the persistence length of dsDNA including electron microscopy [110], atomic force microscopy [111] [112] [113], optical tweezers [114], transient electric birefringence and magnetic beads [115].Though, all the methods adopt different techniques, all of them agreed upon a persistence length of approximately in the range of 45-50nm depending on the method and techniques used to calculate the same.

This flexibility of DNA is not only affected by the local ionic strength but also on the temperature and the solvent in which it lies. Dependence of temperature is of particular interest as DNA bending is known to decrease with increase in temperature. [116] Ionic strength also plays a major role in the stability of the DNA. A simple electrolyte can stabilize the dsDNA by screening the electrostatic repulsion between the phosphate groups and making them more stable. It was seen that the increase in the salt concentration lead to a decrease in the persistence length and this was probably due to the increased shielding of the negative charge of the phosphate backbone by the added counterions. [116] DNA persistence length measured with around 100μM solution of Mg^{2+} ions is known to be independent of the Mg^{2+} ion concentration. Normally Mg^{2+} ions have no global effect on the structure of DNA except for its shielding of the charge of the negative phosphate backbone.

2.5.1 POLYMER EXTENSION MODELS

A polymer chain in the solution continuously goes on changing its shape. The shape of the polymer at any given instant of time is called its conformation. If we pull the two ends of the polymer and it reaches its full extension mode, the length from end to end is called the contour length (Figure 33). The contour length directly depends on the molecular weight of the polymer.

Normally, in solution form, the chain never adopts this fully stretched conformation but is rather in a crumpled coil form and is called random coil and the polymer is supposed to follow a 3D random walk.

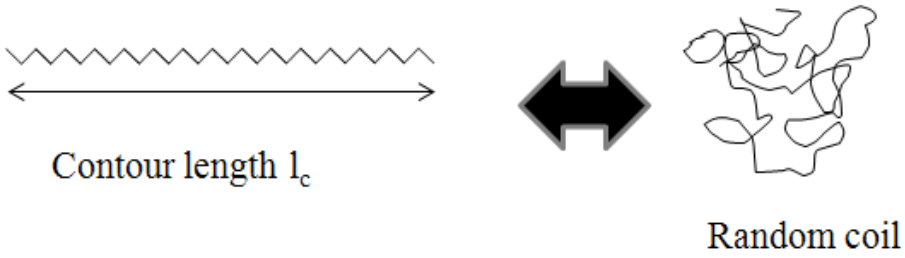


Figure 33. Contour length of a random coil

a. FREELY JOINTED CHAIN MODEL

The simplest model to describe a polymer is the freely jointed chain model (FJC) and the polymer is referred to as an ideal chain. This model assumes the polymer as a random walk and made up of orientationally independent monomers bearing no correlation or interactions among each other to the order that the two monomers can coexist at the same place. In the freely jointed chain model, although the bond length is fixed but the bonds are free to rotate, implying that the subunits are joined by perfectly rigid hinges. for a flexible polymer chain of n+1 atoms, the bond length can be described as $\vec{a}_i$.

The end to end vector is the sum of all the n bond vectors in the polymer chain and is given as:

$$\vec{R} = \vec{R}_n - \vec{R}_0 = \sum_{i=1}^{n} \vec{a}_i$$

This number varies from chain to chain as all chains have their own respective bond vectors. For an isotropic system, this number tends to 0, as all the bond lengths and angles are the same and there is no preference of any direction.

$$\left\langle \vec{R}_n \right\rangle = 0$$

The mean square end to end distance gives an average measure of the size of the polymer and is given as:

$$\left\langle \vec{R}^2 \right\rangle = \left\langle \vec{R}_n{}^2 \right\rangle = \left\langle \vec{R}_n \bullet \vec{R}_n \right\rangle$$

$$\left\langle \left(\sum_{i=1}^{n} \vec{a}_i \right) \bullet \left(\sum_{j=1}^{n} \vec{a}_j \right) \right\rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} \left\langle \vec{a}_i \bullet \vec{a}_j \right\rangle$$

According to the assumption, that all the bond vectors have the same length l,

$$l = \left| \vec{a}_i \right|$$

, the dot product between the bond vectors making an angle can be written as:

$$\left\langle \vec{a}_i \bullet \vec{a}_j \right\rangle = l^2 \cos\theta_{ij}$$

As there are no correlations between the orientations of the bond vectors, this implies that

$$\left\langle \cos\theta_{ij} \right\rangle = 0 \quad \text{for i} \boxtimes \text{j}$$

with only n nonzero terms, where i=j;

$$\left\langle \cos\theta_{ij} \right\rangle = 1 \quad \text{for} \quad \text{i=j}$$

Thus, giving the final value of the mean square end-to -end distance as:

$$\left\langle \vec{R}^2 \right\rangle = l^2 \sum_{i=1,}^{n} \sum_{j=1}^{n} \left\langle \cos\theta_{ij} \right\rangle$$

with

$$\left\langle \vec{R}^2 \right\rangle = nl^2$$

With reference to any bond, the sum over all other bonds converges to a finite number and can be written as:

$$D_i^{'} = \sum_{j=1}^{n} \left\langle \cos\theta_{ij} \right\rangle$$

thereby changing the mean square average end-to-end distance as:

$$\left\langle \vec{R}^2 \right\rangle = l^2 \sum_{i=1}^{n} D_i^{'} = D_n nl^2$$

where D_n is called the Flory's characteristic ratio and can be written as:

$$D_n = \frac{1}{n}\sum_{i=1}^{n} D_i'$$

It is the average of D_i' over all the possible bonds in the polymer and is always greater than 1 for all the polymers. For infinite chains, this number saturates as all the ideal polymers ignore the effects of steric hinderance between the monomers or segments. Thus, for infinite chains, mean square end-to-end distance is written as:

$$\left\langle \vec{R}^2 \right\rangle \cong D_\infty n l^2$$

Ideal polymers can also be described using equivalent freely jointed chain where local interactions are taken into account by including D_∞ term.

Although, it has the same $\left\langle \vec{R}^2 \right\rangle$ and the same end to end distance(max) as in the case of an actual polymer, but it differs in the point that the polymer now ends up having N bonds of length d, called the kuhn length,which are freely joined to each other. The contour length of the polymer can now be written as:

$$Nd = R_{\max}$$

and the mean square end to end distances:

$$\left\langle R^2 \right\rangle = Nd^2 = dR_{\max} = D_\infty n l^2$$

On further simplification of these equations, the number of freely jointed bonds and the kuhn length can be written as [117], [118]:

$$d = \frac{\left\langle R^2 \right\rangle}{R_{\max}} = \frac{D_\infty n l^2}{R_{\max}}$$

$$N = \frac{{R_{\max}}^2}{D_\infty n l^2}$$

The major drawback in using FJC model is that it tends to coarse grain the description of the polymer molecule with the assumption that the kuhn segment has a fixed length and is

completely straight and unstretchable. It does not takes into account any thermal fluctuation away from the straight lines, making the polymers to fluctuate, move or disorder only at the freely oriented joints but this is not the real case scenario of the polymers.

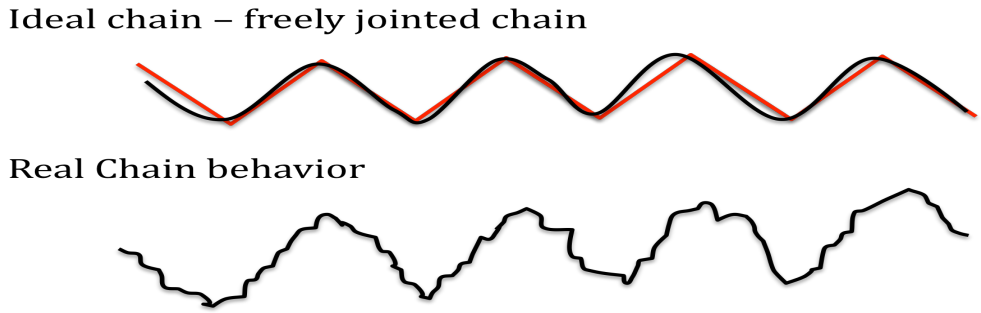


Figure 34. Ideal chain Vs the real chain

Also, though the force measured by this model agrees at very low (<100 fN) and high forces (>5pN), but it fails at the intermediate force Figure 6.

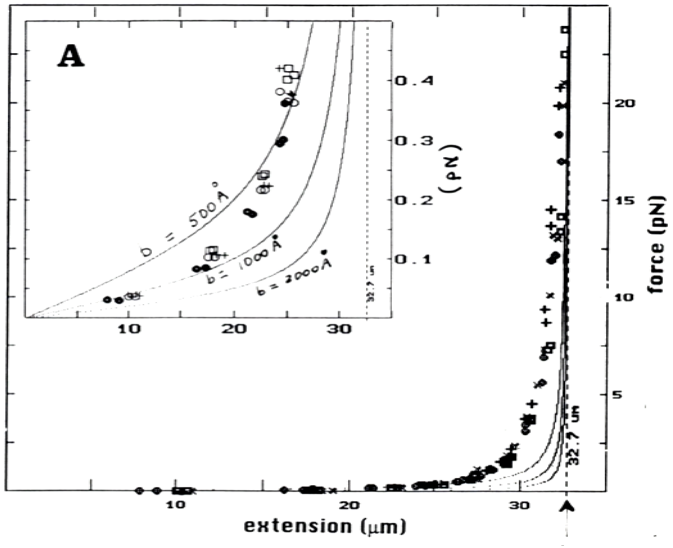


Figure 35. FJC model at low, intermediate and high forces

b. WORM LIKE CHAIN MODEL

This model is sometimes referred to as Kratky –Porod model [109], and is mostly used to describe the behavior of semiflexible polymers whose flexibility increases as their length is increased. It envisions the polymer to be in a continuous elastic medium, displaying a sort of

connectivity or correlation between the successive segments, with all the segments pointing in the same direction. This model is characterized by an exponentially decaying correlation function given as:

$$f(s) = \exp\left(-\frac{s}{l_p}\right)$$

where lp is the persistence length that describes the decay of the correlation with the increase in the length. Coorelation is described in terms of the changes in the direction of the chain along the contour. In this, the bond angle is fixed and is very small. This is a good model for DNA or other semiflexible polymers, as here the flexibility in the polymer is not due to the bond angle but it is due to the fluctuations along the length or contour of the chain from the straight line.

Again, the mean square end to end distance can be written as:

$$\left\langle \vec{R}^2 \right\rangle = \left\langle \vec{a}_i \bullet \vec{a}_j \right\rangle = l^2 \left\langle \cos\theta_{ij} \right\rangle = l^2 \left\langle \cos\theta^{|j-i|} \right\rangle$$

where b is the very short bond length

$$l = b|i-j|$$

$$\cos\theta \cong 1 - \frac{\theta^2}{2}$$

Substituting the values:

$$\left\langle \vec{a}_i \bullet \vec{a}_j \right\rangle = \left[1 - \frac{\theta^2}{2}\right]^{|i-j|} = \left[1 - \frac{\theta^2}{2}\right]^{\frac{l}{b}}$$

$$l_p = \frac{2b}{\theta^2}$$

$$\left\langle \vec{a}_i \bullet \vec{a}_j \right\rangle = \left[1 - \frac{b}{l_p}\right]^{\frac{l}{b}} = \exp\left(\frac{-l}{l_p}\right)$$

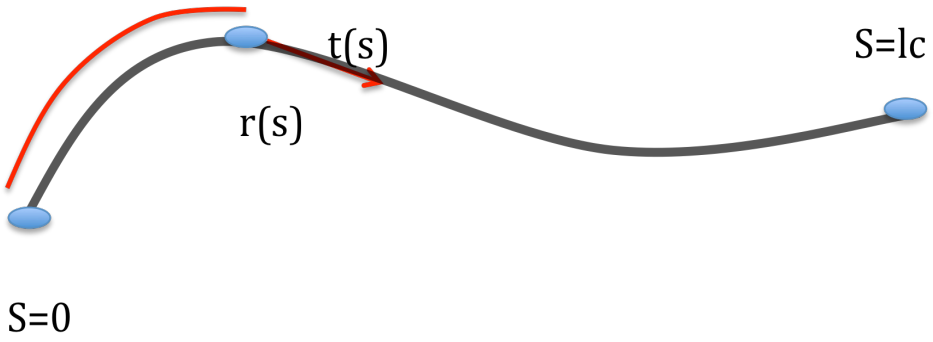


Figure 36. WLC model

In the figure 36, r(s) is the position vector along the contour s, with t(s) as the tangential vector given by:

$$t(s) = \frac{\partial r}{\partial s}$$

and the end to end distance or the contour length can be written as:

$$\vec{R} = \int_0^l r(s)ds$$

$$\langle t(s) \bullet t(0) \rangle = \langle \cos\theta(s) \rangle = \exp(-s/l_p)$$

$$\langle R^2 \rangle = \left\langle \int_0^l t(s)ds \bullet \int_0^l t(s')ds' \right\rangle$$

$$\langle R^2 \rangle = 2l_p l\left[1 - \frac{l_p}{l}\left(1 - \exp(-l/l_p)\right)\right]$$

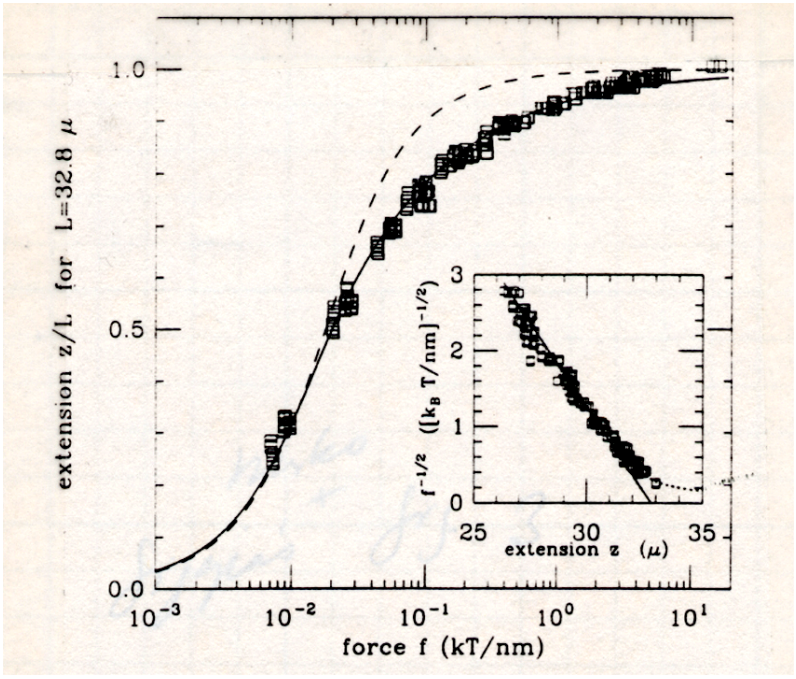


Figure 37 WLC model at low, intermediate and high forces

2.6 PROJECT

In this section, we present a study comparing methylated DNA and normal DNA wrt its persistence length and contour length. Although, previous experiments and studies show nodifference between the physical properties of the two [119], [120], the data collected and interpreted here gives a different picture to the methylation phenomena and its effect on gene silencing.

2.6.1 MATERIALS AND METHODS

2.6.1.1 ISOLATION OF THE DNA FROM E.COLI

E.Coli were transformed to express the required htert-601 strain in their genetic material .The transformed bacteria were cultured in LB (lysogeny) broth on a large scale to yield the plasmid. Amplicillin antibiotic was added to the LB broth while culturing to prevent the growth of other strains of the bacteria. Plasmid (6991 bp) was purified out of the bacterial pellet using Qiagen Giga prep kit and was further subjected to double digestion by the restriction enzymes BamH1 and Ecor1, releasing out 2905 bp of the plasmid DNA containing htert- 601 sequences. For the methylation studies the purified plasmid was treated with CpG Methyltransferase (M.SssI)

and S-Adenosine methionine (SAM) as a cofactor for the reaction. The methylation reaction was preceded overnight and the products were further purified to give methylated DNA. Changing the amount of CpG methyltransferase and SAM produced various levels of methylated DNA. Incubating the mDNA and the normal DNA with the restriction enzyme Ava1 for 1 hr tested methylation. DNA were purified by running them through gel electrophoresis and further purifying and extracting DNA out of these gels and was further dissolved in 15mM NaCl and 1mM phosphate solution. The purified fractions of the DNA were imaged in AFM.

2.6.1.2 AFM IMAGING

The sample prep was done on a freshly cleaved mica surface. Solution of 10mM Mg^{++} was prepared and added to the mica surface, immediately followed by DNA at a concentration of 1 ng/μl. The sample was incubated on the mica surface for 2 min and was washed gently with distilled water and dried with nitrogen. The imaging was done on AFM 5500 from Agilent Technologies using Si_3N_4 cantilevers with spring constant ranging from 25- 75N/m and resonating around 300 kHz. The results for the persistence length and contour length were derived using worm like chain fits. (Matlab code in Appendix)

2.6.2 RESULTS AND DISCUSSION:

1. Isolation of htert promoter -601 sequence from the E.Coli pellet.

The plasmid obtained from the E.Coli pellet is as Figure 38.

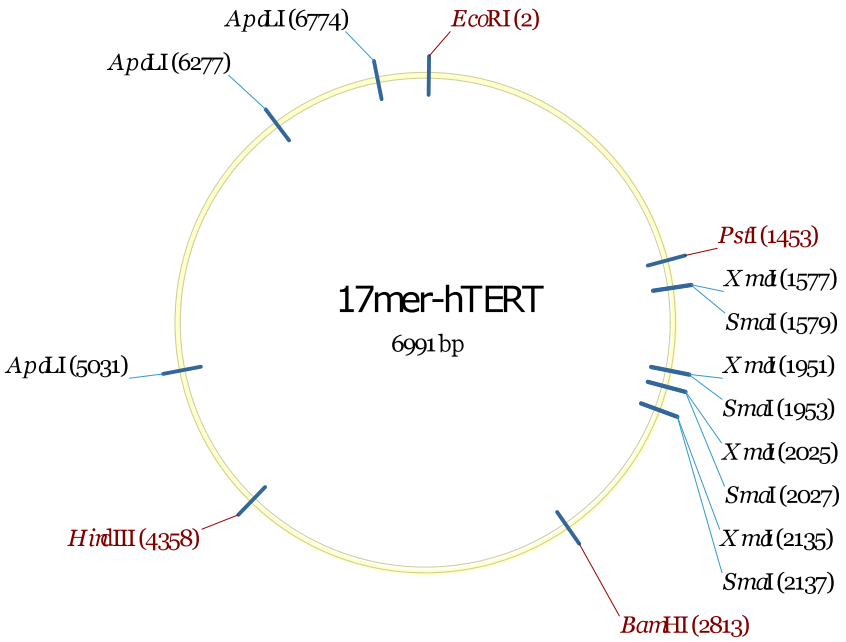


Figure 38. 17mer-htert plasmid

Htert promoter -601 sequence was obtained by double digesting the obtained plasmid as in Figure38 with Ecor1 (site of action-5'….GAATTC….3') and BamH1 (site of action -5'….GGATCC….3'), followed by gel purification. The obtained DNA sequence to be used in all the studies henceforth is as seen in Figure 39.

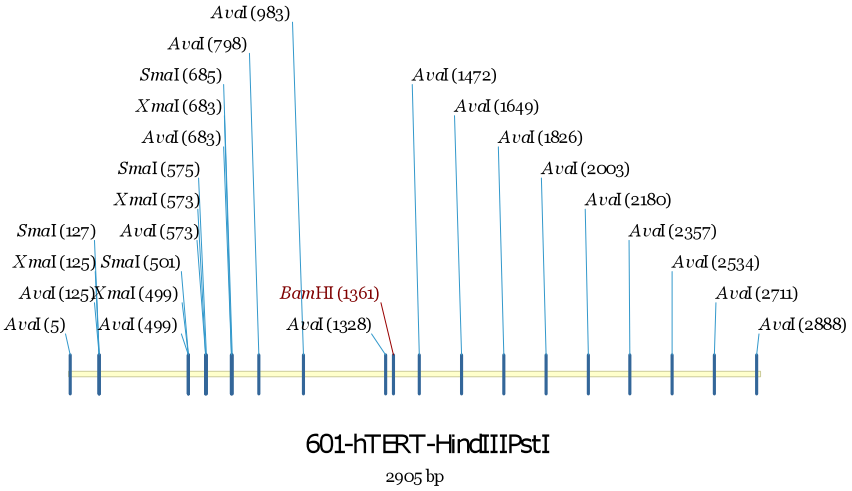


Figure 39. 601-htert promoter sequence with the sites of action of the restriction enzymes

htert promoter 601 repeats 30bp linker DNA

gcactcgggccaccagctccttcaggcaggacacctgcgggggaagcgccctgagtcgcctgcgctgctctccgcatgtcgctggttcccc

ccggccgccctcaaccccagccggacgccgaccccggggaggcccacctggcggaaggaggggcggcgggggcggccgtgcgt

cccagggcacgcacaccaggcactgggccaccagcgcgcggaaagccgccgggtccccgcgctgcaccagccgccagccctggggcc

ccaggcgccgcacgaacgtggccagcggcagcacctcgcggtagtggctgcgcagcagggagcgcacggctcggcagcggggagcg

cgcggcatcgcgggggtggccggggccagggcttcccacgtgcgcagcaggacgcagcgctgcctgaaactcgcgccgcgaggagag

ggcggggccgcggaaaggaaggggagggggctgggagggcccggaggggggctgggccggggacccgggaggggtcgggacgggg

cggggtccgcgcggaggaggcggagctggaaggtgaaggggcaggacgggtgcccgggtccccagtccctccgccacgtgggaagcg

cggtcctgggcgtctgtgcccgcgaatccactgggagcccggcctggccccgacagcgcagctgctccgggcggacccgggggtctggg

ccgcgcttccccgcccgcgcgccgctcgcgctcccagggtgcagggacgccagcgagggccccagcggagagaggtcgaatcggccta

ggctgtggggtaacccgagggaggggccatgatgtggaggccctgggaacaggtgcgtgcggcgacccctttggccgctggcctgatccg

gagacccagggctgcctccaggtccggacgcggggcgtcgggctccgggcaccacgaatgccggacgtgaaggggaggacggaggc

gcgtagacgcggctggggacgaacccgaggacgcattgctccctggacgggcacgcgggacctcccggagtgcctccctgcaacacttc

cccgcgacttgggctccttgacacaggcccgtcatttctctttgcaggttctcaggcggcgaggggtccccaccatgagcaaaccaccccaaa

tctgttaatcacccaccggggcggtcccgtcgagaaagggtgggaaatggagccaggcgctcctgctggccgcgcaccgggcgcctcaca

ccagccacaacggccttgaccctgggccccggcactctgtctggcagatgaggccaacatctggtcacatcccgcccgcacagggtggag

ggcaacctcggggtccaggcacctggctccaagcctcgtggatccgcttgatcgaacgtacgcgctgtcccccgcgttttaaacgccaaggg

gattactccctagtctccaggcacgtgtcagatatatacatcctgtcggaccgagctcctcgggatgcatcccgccctggagaatcttggtgcc

gaagccgctcaattggtcgtagcaagctctaccaccgcttaaacgcacgtaagggctgtcccccgcgttttaaccgccaagaggattactccc

cagtctccaggcacgcgtcagatatatacatcctgtgcatgtattgaactcgggatgcatcccgcccctggagaatcttggtgccgaagccgctc

aattggtcgtagcaagctctaccaccgcttaaacgcacgtaagggctgtcccccgcgttttaaccgccaagaggattactccccagtctccagg

cacgcgtcagatatatacatcctgtgcatgtattgaactcgggatgcatcccgcccctggagaatcttggtgccgaagccgctcaattggtcgtag

caagctctaccaccgcttaaacgcacgtaagggctgtcccccgcgttttaaccgccaagaggattactccccagtctccaggcacgcgtcaga

tatatacatcctgtgcatgtattgaactcgggatgcatcccgcccctggagaatcttggtgccgaagccgctcaattggtcgtagcaagctctacc

accgcttaaacgcacgtaagggctgtcccccgcgttttaaccgccaagaggattactccccagtctccaggcacgcgtcagatatatacatcct

gtgcatgtattgaactcgggatgcatcccgcccctggagaatcttggtgccgaagccgctcaattggtcgtagcaagctctaccaccgcttaaac

gcacgtaagggctgtcccccgcgttttaaccgccaagaggattactccccagtctccaggcacgcgtcagatatatacatcctgtgcatgtattg

aactcgggatgcatcccgcccctggagaatcttggtgccgaagccgctcaattggtcgtagcaagctctaccaccgcttaaacgcacgtaagg
gctgtcccccgcgttttaaccgccaagaggattactccccagtctccaggcacgcgtcagatatatacatcctgtgcatgtattgaactcgggat
gcatcccgcccctggagaatcttggtgccgaagccgctcaattggtcgtagcaagctctaccaccgcttaaacgcacgtaagggctgtccccc
gcgttttaaccgccaagaggattactccccagtctccaggcacgcgtcagatatatacatcctgtgcatgtattgaactcgggatgcatcccgcc
ctggagaatcttggtgccgaagccgctcaattggtcgtagcaagctctaccaccgcttaaacgcacgtaagggctgtccccgcgttttaaccg
ccaagaggattactccccagtctccaggcacgcgtcagatatatacatcctgtgcatgtattgaactcgggagatctgcatgca

As seen in the sequence, the DNA is rich in CpG content which is a characteristic of the promoter sequence and thus methylation can be studied on a broader basis with cytosine being the sites of methylation. The sequence has 64% total GC content with htert promoter sequence with 1489 bp followed 9 – 601 repeats separated by 30bp linker DNA.

m-DNA and DNA samples were suspended in 1/10 PBS (15mM NaCl, 1mM phosphate buffer) and were further diluted at a concentration of 1 ng/µl. Methylation test was done with Ava1 (Figure 40) enzyme (site of action - 5'…. CYCGRG….3' , where Y can be C or T and R can be A or G) and since methyl group on getting methylated by CpG methyltransferase protects the CpGs, the restriction enzyme was not able to cut the mDNA but was able to do so with normal DNA as seen in gel in Figure 40.

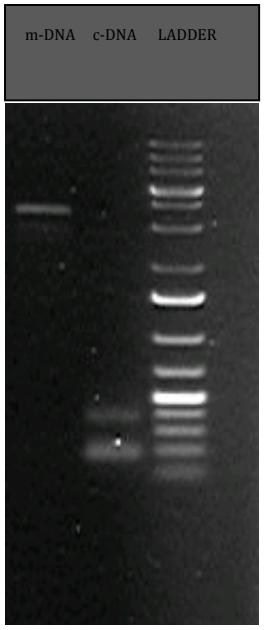


Figure 40. Methylation test using Ava1 enzyme.

AFM images of the control DNA and the mDNA were taken and are as shown in figure 41.

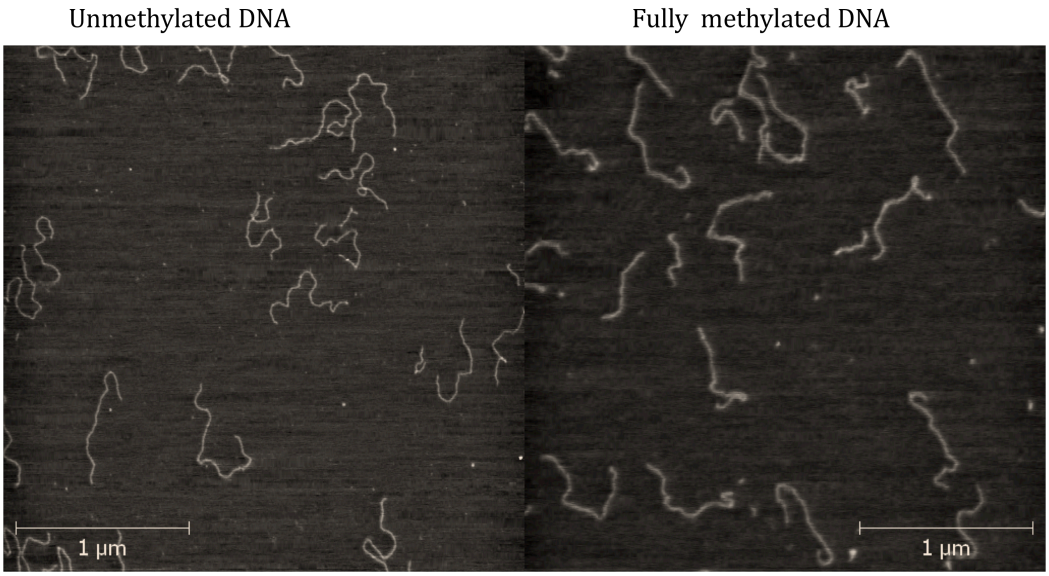


Figure 41. AFM images of control DNA and methylated DNA

Persistence length and contour length measurements were carried out on the individual DNA strands using the MATLAB program code given in the Appendix and the data was analyzed and plotted as seen in Figure 42. More than 100 DNA molecules were taken for each measurement and fitted with worm like chain model to give the persistence length and the contour length. For all the molecules, it was made sure that they were equilibrated by measuring the mean square separation of pairs of points on the same DNA strand. This is given in the Matlab code in the Appendix. For the calculation of persistence length, tangent -tangent correlation function between any two points on the DNA strand was calculated for the length of the DNA and this correlation for an equilibrated DNA must be approximately equal to the orientational correlation function described in equation 2(Page 54), thus giving the measure of the persistence length.[121]

Multiple runs of both methylated DNA and the control DNA were done to ensure the reproducibility of the result as no drastic changes in persistence length as well contour length have been reported in literature till date.

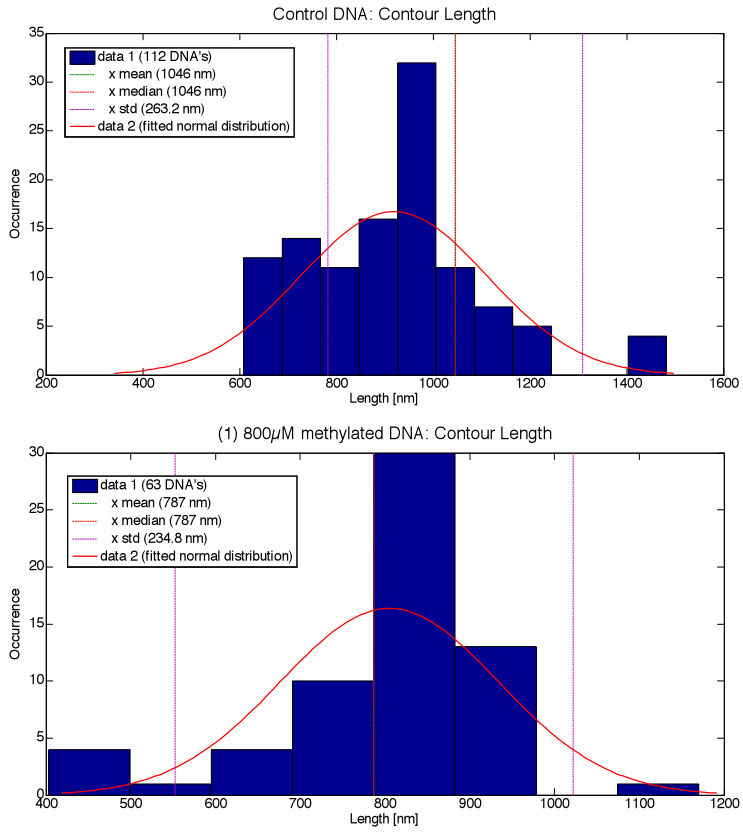


Figure 42.Contour length of methylated DNA and control DNA

Sample run	Samples	Persistence length(nm)	Contour length(nm)
1.	Control DNA	47.0 ± 9.5	1046.0 ± 263.2
2.	Control DNA	53.8 ± 3.23	1023.9 ± 112.3
3.	Control DNA	50.7 ± 4.23	1033.9 ± 102.3
1.	Methylated DNA	92.6 ± 11.5	783.0 ± 234.8
2.	Methylated DNA	92.0 ± 5.0	828.6 ± 92.3
3.	Methylated DNA	92.1 ± 4.0	838.6 ± 72.2

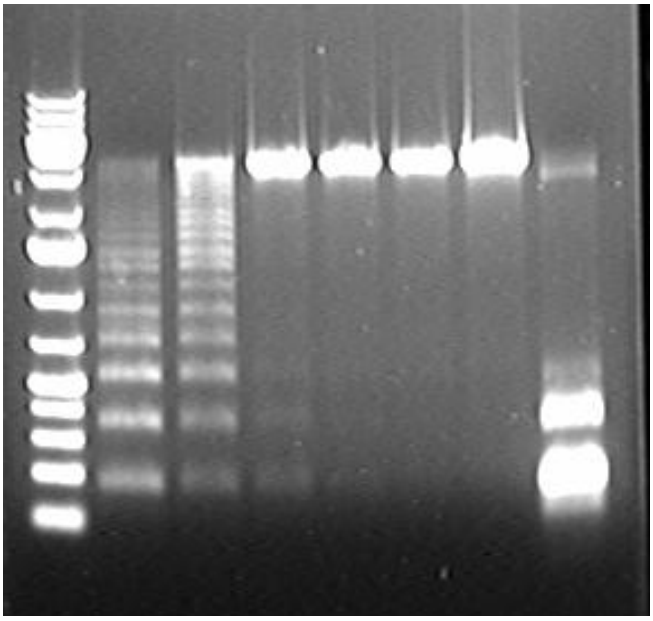
Table 4 Various runs of methylated and control DNA samples

After getting these numbers, the next step was to evaluate if the lower levels of methylated DNA also exhibited the same trend. Varying the amount of SAM, which is the methyl group donor, we were able to get various levels of methylated DNA. To know how much methylated the DNA were, intensity patterns of the gels were taken into account from the gel analyses software and were plotted. As know that the fully methylated DNA cannot be cleaved by Ava1 Enzyme, but as the levels of methylation is reduced, more and more free CpG groups are exposed and this results in a ladder of small DNA's as seen in figure 43. The level of methylation is analyzed from this gel using the program in Appendix 2.

Since there are 17 sites of action of Ava1 enzyme on the DNA template, the data is plotted as to how the methylation patterns change when the sites are getting blocked by methylation and thereby are protected. Comparing the plots from the program and the intensity data from the gels, we were able to establish the level of methylation.

GEL 1

5µM 20µM 60µM 100µM 200µM 800µM Control



GEL 2 control 100uM 60uM 20uM 5uM

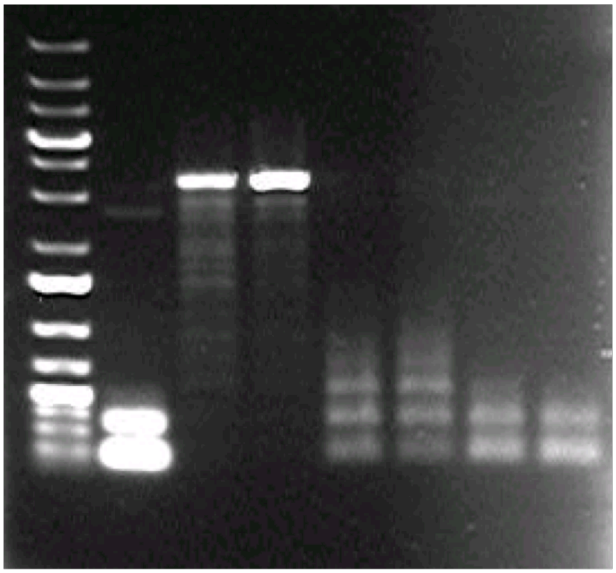
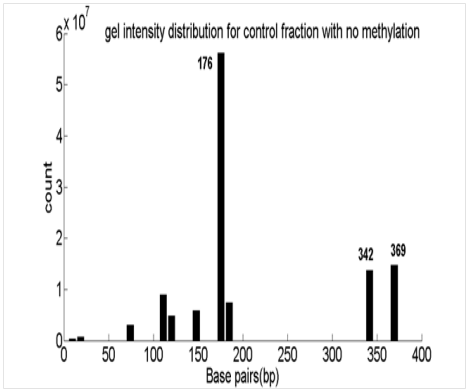
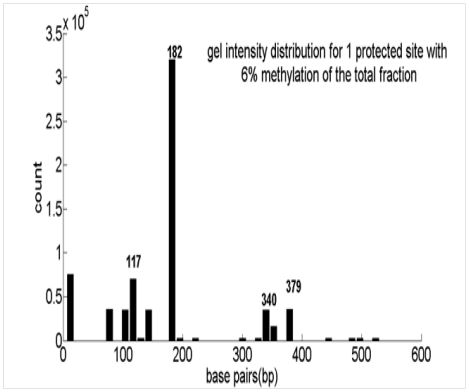
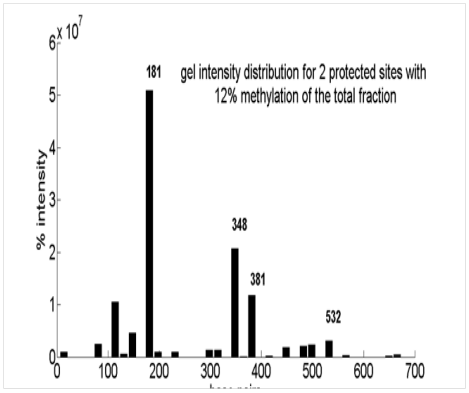


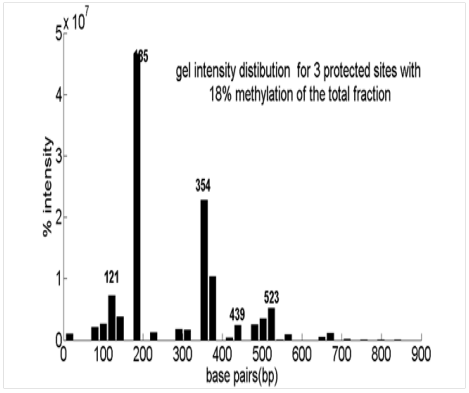
Figure 43 Electrophoresis gel showing action of Ava1 enzyme on the various levels of methylated DNA

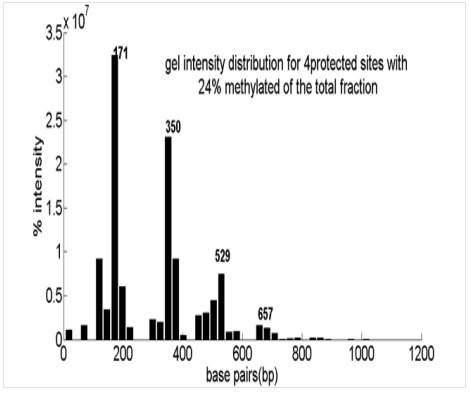
Figure 44. Plots for all protected sites

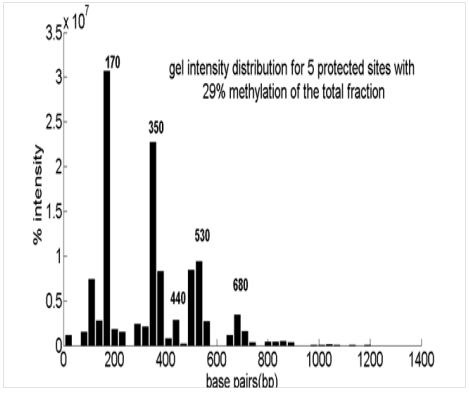


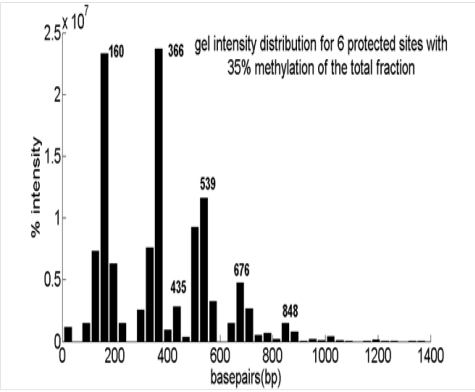


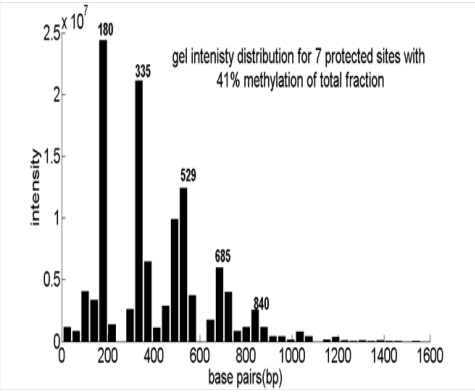


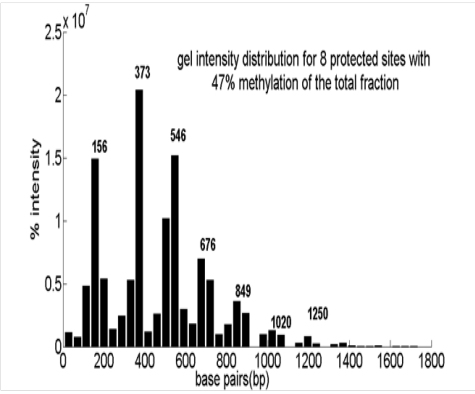


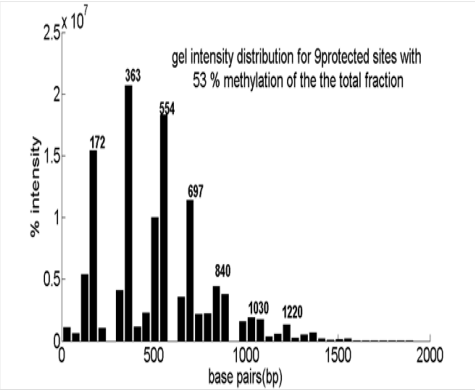


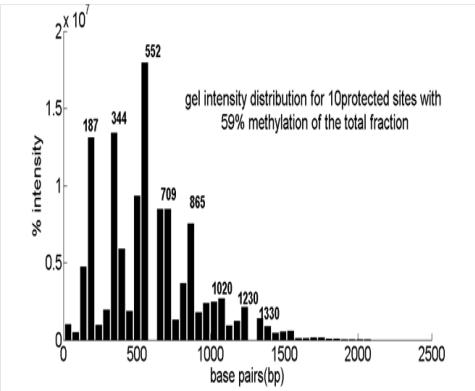


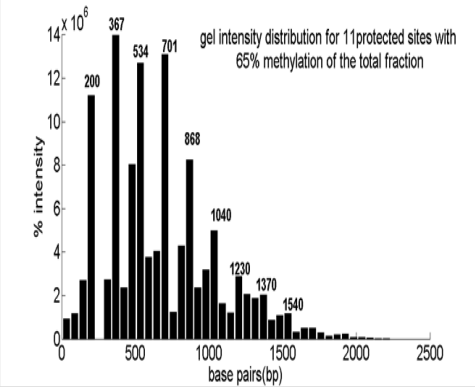

gel intensity distribution for 6 protected sites with
35% methylation of the total fraction
x 10^7
% intensity
basepairs(bp)
160
366
435
539
676
848

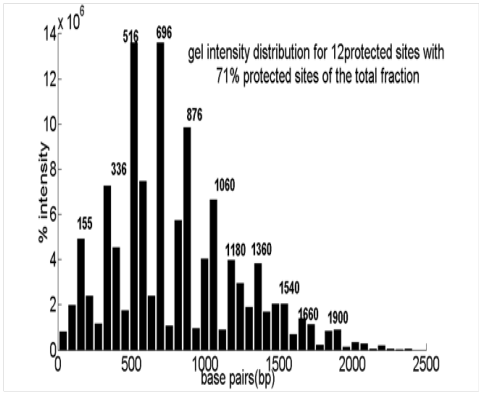

gel intenisty distribution for 7 protected sites with
41% methylation of total fraction
x 10^7
intensity
base pairs(bp)
180
335
529
685
840

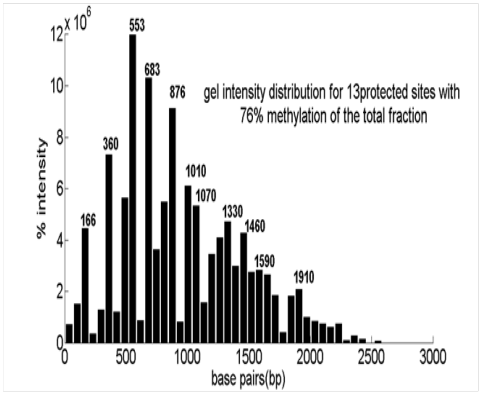

gel intensity distribution for 8 protected sites with
47% methylation of the total fraction
x 10^7
% intensity
base pairs(bp)
156
373
546
676
849
1020
1250

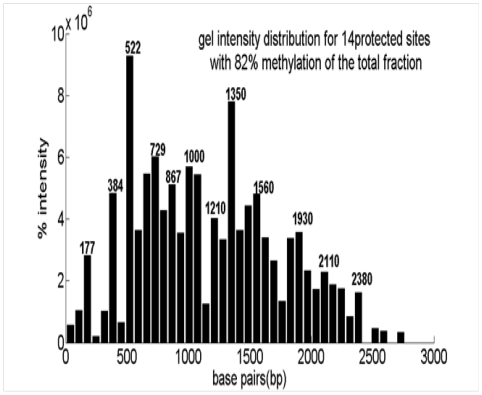

gel intensity distribution for 9protected sites with
53 % methylation of the the total fraction
x 10^7
% intensity
base pairs(bp)
172
363
554
697
840
1030
1220

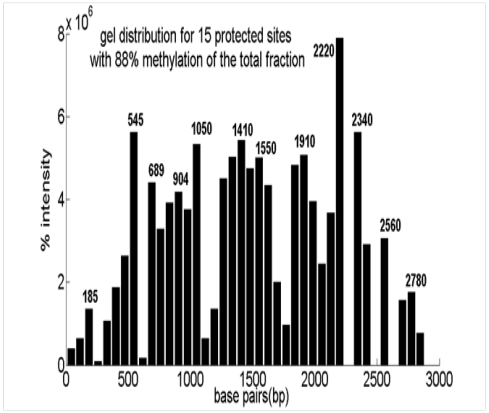

gel intensity distribution for 10protected sites with
59% methylation of the total fraction
x 10^7
% intensity
base pairs(bp)
187
344
552
709
865
1020
1230
1330

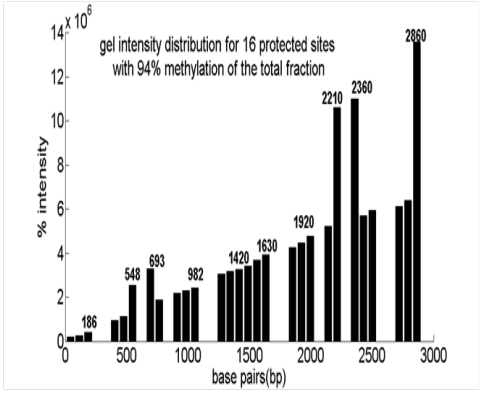

gel intensity distribution for 11protected sites with
65% methylation of the total fraction
x 10^6
% intensity
base pairs(bp)
200
367
534
701
868
1040
1230
1370
1540

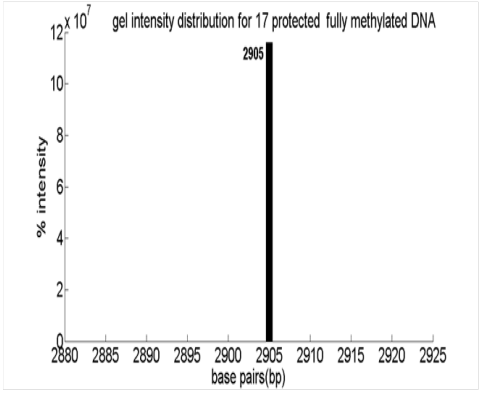

x 10^6
gel intensity distribution for 12protected sites with
71% protected sites of the total fraction
% intensity
base pairs(bp)
155
336
516
696
876
1060
1180
1360
1540
1660
1900


x 10^6
gel intensity distribution for 13protected sites with
76% methylation of the total fraction
% intensity
base pairs(bp)
166
360
553
683
876
1010
1070
1330
1460
1590
1910


x 10^6
gel intensity distribution for 14protected sites
with 82% methylation of the total fraction
% intensity
base pairs(bp)
177
384
522
729
867
1000
1210
1350
1560
1930
2110
2380


x 10^6
gel distribution for 15 protected sites
with 88% methylation of the total fraction
% intensity
base pairs(bp)
185
545
689
904
1050
1410
1550
1910
2220
2340
2560
2780


x 10^6
gel intensity distribution for 16 protected sites
with 94% methylation of the total fraction
% intensity
base pairs(bp)
186
548
693
982
1420
1630
1920
2210
2360
2860


x 10^7
gel intensity distribution for 17 protected fully methylated DNA
% intensity
base pairs(bp)
2905

Lanes:	5uM		20uM		60uM		100uM		400uM		800uM		Control
Bands	bp	%	bp	%	bp	%	bp	%	bp	%	bp	%	bp
1	3000	11.1	3000	20	3000	64.8	3000	63.2	3000	100	3000	100	895
2	3000-2000	9.04		4.2	483	16.4	414	20.6					514
3	2000-1500	4.19		8.5	245	18.7	240	16.0					346
4			2417	5.0									
5	1842	5.4	2000	5.7									
6	1629	5.6	1842	5.2									
7	1377	6.8	1651	5.2									
8	1066	10.3	1377	7.1									
9	765	11.4	1089	7.0									
10	467	13.4	765	9.6									
11	255	17.0	467	10.5									
12			245	11.6									
Sum		100		100		100		100		100		100	
In Lane		100		100		100		100		100		100	

Table 5 for the methylation intensity from the gel 1

Lanes:	100uM		60uM		20uM		5uM		Control	
Bands	**bp**	**%**	**bp**	**%**	**bp**	**%**	**bp**	**%**	**bp**	**%**
1	3000	48	3000	54	3000	2.039	885	19.461	3000	5.4042
2	2896	17	2896	15	1500	4.01	708	26.502	300	42.496
3	1959	15	2223	11	1225	5.7472	531	15.02	122	52.1
4	1470	5	2000	5.5	1084	8.724	354	18.782		
5	1107	2.447	1805	4.3	924	6.761	177	20.235		
6	906	2.15	1500	2.92	885	16.098				
7	773	3.3141	1328	1.59	700	6.1351				
8	700	3	1041	1.35	555	19.073				
9	555	2.6	980	1.3	339	15.209				
10	354	1.5	820	1.5	177	16.252				
11			555	1.54						
12										
Lanes:		100		100		100		100		100
total		100		100		100		100		100

Table 6 for the methylation intensity of gel2

No. of protected sites	Degree of methylation(%)(p)	% CpG methylated fraction (p*64%)
1	5.88	0.04
2	11.76	0.08
3	17.65	0.11
4	23.53	0.15
5	29.41	0.19
6	35.29	0.23
7	41.18	0.26
8	47.06	0.30
9	52.94	0.34
10	58.82	0.38
11	64.71	0.41
12	70.59	0.45
13	76.47	0.49
14	82.35	0.53
15	88.24	0.56
16	94.12	0.60
17	100.00	0.64

Table7 . For % methylation corresponding to the sample

Comparing the values from the Table 5, 6, 7 for the gel intensity and the plots for the protected sites, we can get the values of the % methylation of the individual samples in the gel. Table 8 gives the % methylation for the different samples being imaged. Persistence length and the contour length were measured (using worm like chain fits) for all these samples and the trend for the change in both of them can be seen as in figure 45.

	Conc. SAM	No. of protected sites	Degree of methyl(p)	%CpG methylated (p*64%)
gel1	5 μM	14	0.82	0.53
	20 μM	15	0.88	0.56
	60 μM	16	0.94	0.6
	100 μM	16	0.94	0.6
	800 μM	17	1	0.64
gel2	5 μM	12	0.71	0.45
	20 μM	13	0.76	0.49
	60 μM	16	0.94	0.6
	100 μM	16	0.94	0.6
	800 μM	17	1	0.64

Table 8 Identification of methylation of the samples

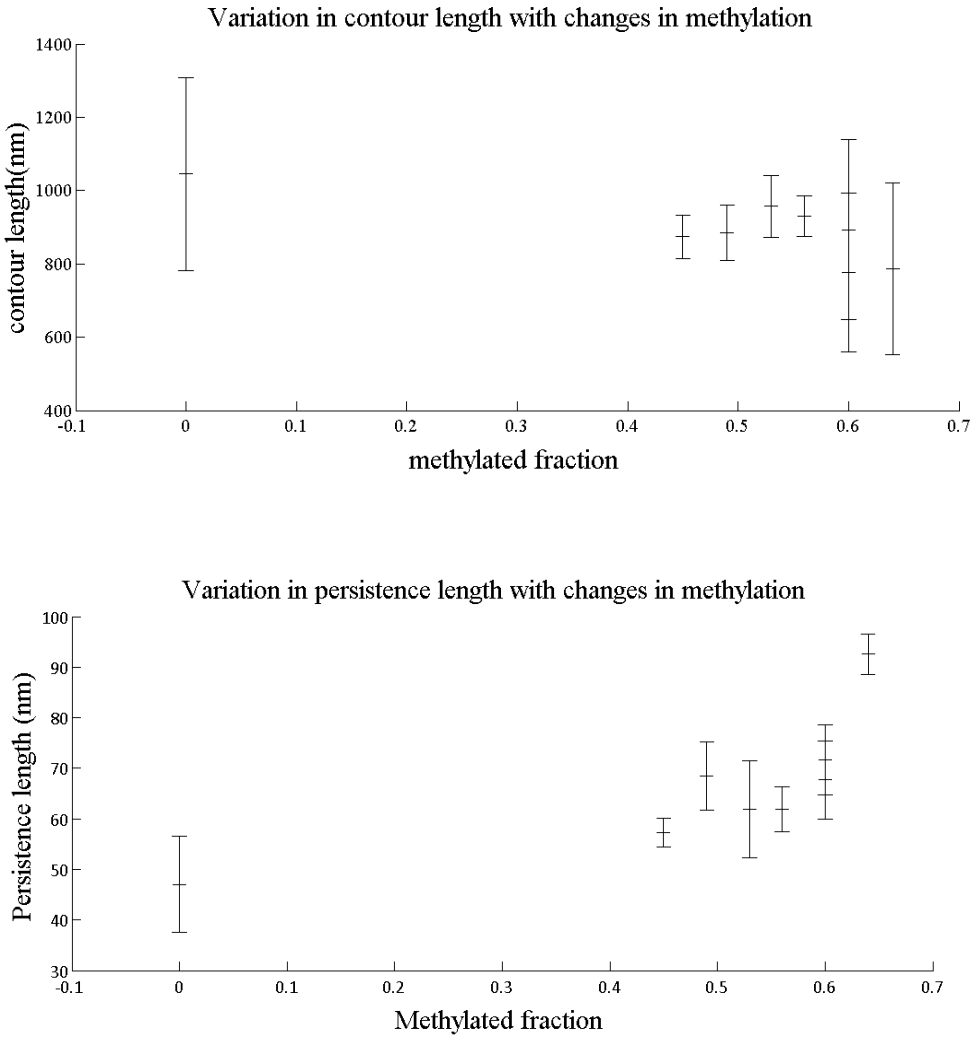


Figure 45 .Variation in contour length with changes in methylation and Variation in persistence length with changes in methylation.(double error signal on the methylated fraction at 0.6 is because of the overlap of two points)

The results are summarized in the Table 9.

	Conc. SAM	% CpG methylated (p*64%)	Persistence length (nm)	Error	Contour length (nm)	Error
gel1	0 μM	0	47	9.5	1046	263
	5 μM	0.53	61.8	9.6	956.9	83.9
	20 μM	0.56	61.9	4.4	931	55.5
	60 μM	0.6	67.7	7.7	893.7	244.6
	100 μM	0.6	71.6	6.9	776	217
	800 μM	0.64	92.5	4.0	787	234.8
gel2	5 μM	0.45	57.3	2.9	873.9	58.8
	20 μM	0.49	68.5	6.7	884.7	75.7
	60 μM	0.6	66.6	11.2	977	62.0
	100 μM	0.6	70.7	7.3	976	82.5
	800 μM	0.64	92.5	11.5	838.6	72.3

Table 9. Results

As seen from the Table 9 and figure 44, it is evident that the methylated DNA behaves very differently as compared to the normal DNA template and these results are very reproducible, leaving any doubts about the experimental errors. This is in sharp contrast to the earlier studies on methylated DNA persistence length measurements. Not only is the persistence length increasing, measured almost double, for fully methylated fraction but the contour length is also decreasing. From the contour length plot in figure44, few observations can be pointed out.

1. Contour length for the control DNA is 1046 ± 263 nm, which agrees with the expected length of this DNA fragment. (2905 bp x 0.34 nm/bp = 987.7 nm). Thus, establishing a control that the experimental setup is working correctly.

2. 0.34 nm/bp for the control DNA , establishes that the control DNA template is a B form DNA.

3. The approximate value for the nm/bp for completely methylated fraction gives a value of 0.27 ± 0.08 nm/bp , indicating that the fully methylated DNA fraction is the A form of DNA and is dehydrated more locally.

If the data from previous studies that the solution persistence length is unchanged by methylation is true, then the gradual increase of persistence length seen from the plot is clearly an interfacial effect. To make sure that what we are observing is an interfacial effect only, control experiments were done to compare the behavior of methylated and control DNA in solution using dynamic light scattering.[123] Both the samples behave in the same way and had an almost equal diffusion coefficients, confirming the fact that the methylated DNA and the control DNA do behave differently at the interface of Magnesium – mica surface, making it an interfacial effect rather than their inherent property. Figure 45 b shows the autocorrelation functions of both methylated and control DNA.

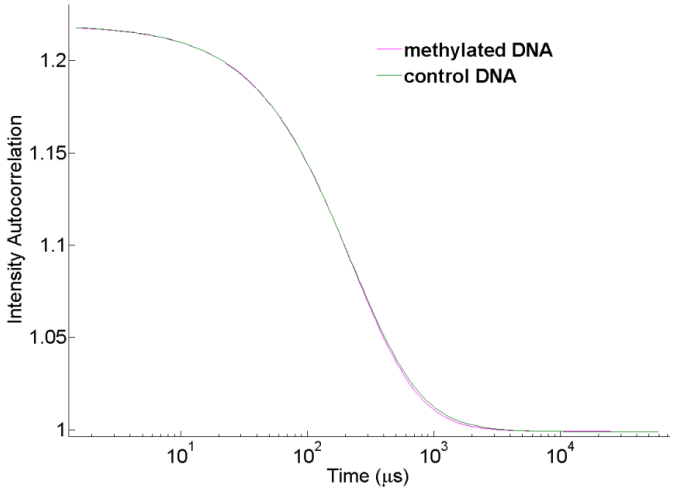


Figure 45b. Autocorrelation function for methylated and control DNA

During the process of methylation, when a proton is replaced by methyl group in cytosine base, the methylated DNA can be seen as a surfactant molecule, having a water repelling hydrophobic part as the methyl group and phosphate backbone as the hydrophilic part. Thus, the mDNA can be viewed as a surfactant. Nothing is more hydrophobic than water . Thus, when the mDNA is added on the Mg- mica surface, which is obviously less hydrophobic than water, the water being more hydrophobic, pushes the methyl groups on the DNA , onto the Mg-mica surface which is less hydrophilic. In simple words, methyl groups will be pushed (or will move) towards the region of lesser hydrophobicity, i.e. towards Mg^{++}.

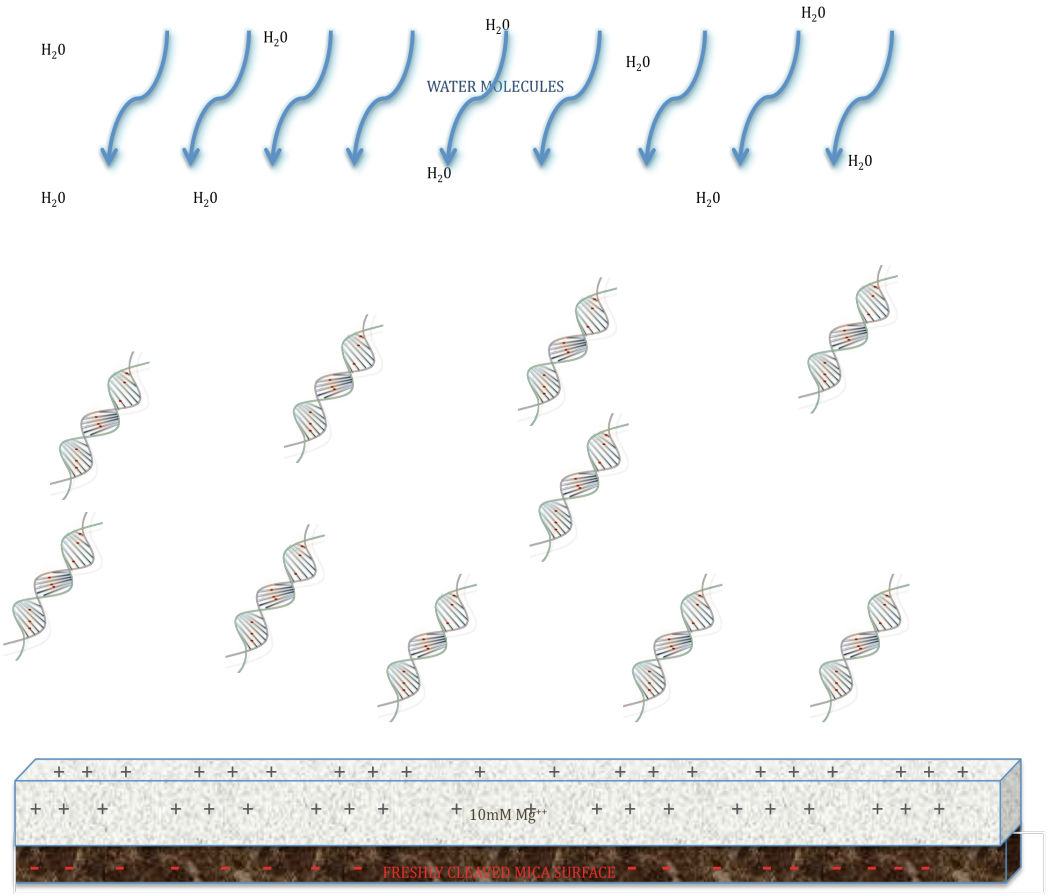


Figure 46. Mechanism of gene silencing by DNA methylation.

But what is the importance or significance of this?

This can be compared to the condition in the genome and can explain the concept of gene silencing when the DNA gets methylated. The DNA, when gets methylated, is pushed by the

water environment all around it in the nucleus of the cell, towards the histone proteins around which they are wrapped. The positively charged histone proteins can be correlated to the Mg^{++} surface. Thus, as the methyl groups on the DNA are getting pushed by water, in the process, the DNA is also pushed towards the histone proteins. The regions of the DNA that are pushed towards the histones become so tightly bound that they can no longer be expressed. This suppression of the genes can have serious consequences on various life processes like cell division, cell growth etc.

Supposedly, the genes are coding for the proteins that are required to stop the cell division process. Thus, with the genes not encoding for the protein, the protein will not be expressed, leading to an uncontrolled pathway for the cell division process which will most likely result in tumor or cancer. Thus, this simple explanation gives a new picture on the process of gene silencing by methylation.

2.6.3 FUTURE WORK

In this section, we were able to see how the action of water pushing on to the DNA can explain the concept of gene silencing due to methylation. It will be very interesting to change the solvent and observe what happens. Nothing is more hydrophilic than water. Thus, it will be more informative to see the effect of a hydrophobic solvent on these interfacial interactions and compare the persistence length in both cases.

2.7 DNA IN DMSO

As discussed in section 2.6.2, water being more hydrophilic ,pushes the DNA towards the Mg^{++}- mica surface, which in turn increases the persistence length. It would be interesting to see what happens if there is a change in the environment of the DNA from the native water. DMSO (Dimethyl sulfoxide) was taken as the solvent to resuspend DNA. DMSO can be used as a solvent, as it can form hydrogen bonding with water molecules and DNA is also stable in it. The partial negative charge on the oxygen atom in DMSO, favors the formation of hydrogen bonding with water. It has also been reported that DNA is stable in up to 25% DMSO solution in water, above

which the DMSO interferes with the hydrogen bonding between the base pairs and then denatures the DNA. Table 10 compares the physical properties of DMSO and water.

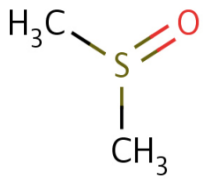


Figure 47 Dimethyl sulfoxide

Solvent	Chemical formula	Boiling point	Dielectric constant	Density	Dispersion	Polar	Hydrogen bonding
Dimethyl sulfoxide (DMSO)	CH_3-S(=O)-CH_3	189 °C	46.7	1.092 g/ml	18.4	16.4	10.2
Water	H-O-H	100 °C	80	1.000 g/ml	15.5	16.0	42.3

Table 10. Comparison of properties of DMSO and water

Being a polar, aprotic solvent, it is difficult for DMSO to donate a H^+ ion, and thus is a hydrogen bond acceptor and not a donor. Also, DMSO-water interaction is stronger as compared to water-water interaction. The average number of water-water hydrogen bonds goes on decreasing on addition of increasing amounts of DMSO, thus signaling a disruption of hydrogen bonding among water molecules. DMSO interaction with water molecule involves 3-H bonds with the molecule, where the oxygen of the DMSO acts as the proton acceptor. This H-bonding involving DMSO molecules leads to a decrease in the H-bonds between the water molecules. The H-bond comprises of a proton of a water molecule and a lone pair in the valence shell of the oxygen atom in any of the molecule, either water or DMSO. The oxygen atom has two lone pairs of electrons in water, and DMSO formulated as a resonance hybrid, has varied 2 or 3 number of

lone electron pairs in the valence shell of the oxygen atom in oxygen. All these properties with H_2O molecule make DMSO an ideal solvent of interest for our studies with DNA. [124]

2.7.1 MATERIALS AND METHODS

Control DNA (unmethylated) and fully methylated DNA were taken as a DNA template for the experiment. The templates were suspended in 5% DMSO. Mica was cleaved freshly and 2µl of 10mM Mg^{++} was added on it followed by addition of DNA (in 5% DMSO). The sample was incubated for 2-3 min and was later washed with fresh distilled water and dried with nitrogen. These samples were further imaged in AAC mode in AFM 5500 with cantilevers from AppNano SPM with a spring constant ranging from 25-75 N/m and resonance frequency around 300 kHz.

2.7.2 RESULTS AND DISCUSSION

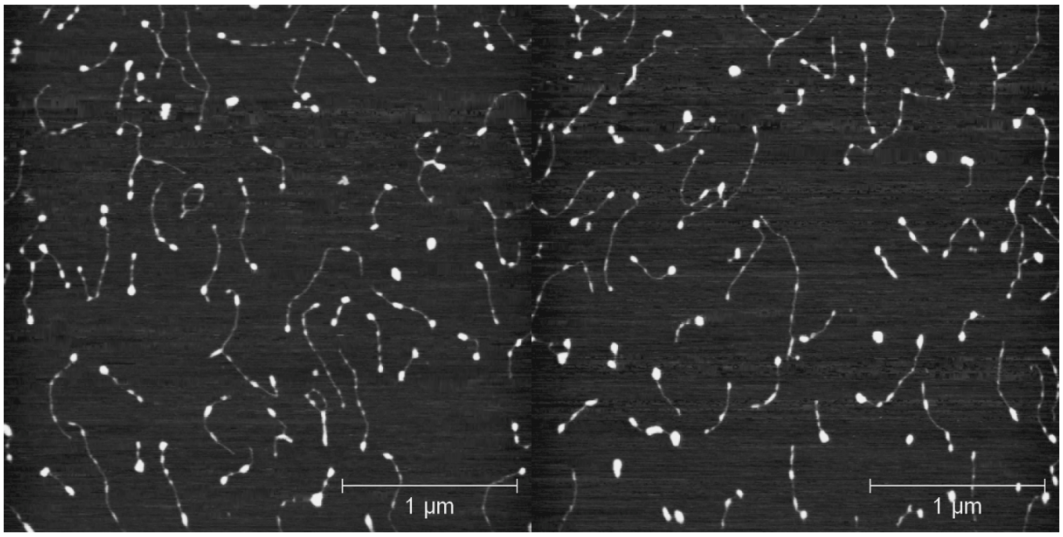


Figure 48 Control DNA in 5% DMSO

From these images, there are few observations very apparent.

1. The DNA molecule is no longer uniform cylindrical polymer molecules, but they have changed their appearance to beads on a string like.

2. All the DNA molecules have big spherical blobs at their ends. In addition to these spherical blobs at the end, there are small sized spherical blobs towards the inside of the polymers.

3. From the images, it looks like the blobs are formed in not just a single DNA molecule, but also few DNA molecules are combining with each other.

On repeating the same experiment on methylated DNA, we get a completely different picture as seen in Figure 49. Methylated DNA no longer has extended polymer structure but is converted to a big spherical blob.

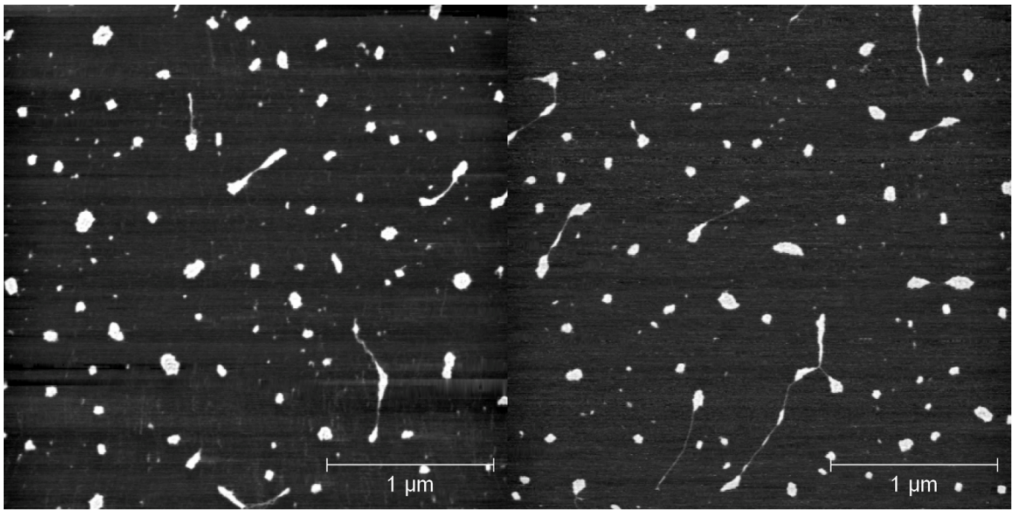


Figure 49. Methylated DNA in 5% DMSO

As seen from the images the DNA molecules look completely different. Control DNA gives like a beads on string appearance and the methylated DNA gives a completely globular appearance. Also individual molecules are connected to each other. From these images and observations already taken into account, it will be interesting to do some data analysis and study on the variation of the contour length, diameter and height of these blobs wrt the number of the blobs for control and methylated DNA.

Table 11 gives an estimate on the frequency of the blobs. The highest number of DNA molecules has two blobs, mostly at both of their ends, followed by 1 at one of the ends.

Data Analysis for Control DNA:

No. of DNA blobs	No. of molecules	Total %
1	48	24.2
2	**93**	**47**
3	41	21
4	11	5.6
5	4	2
6	0	0
7	1	0.5

Table 11. Frequency of the number of blobs

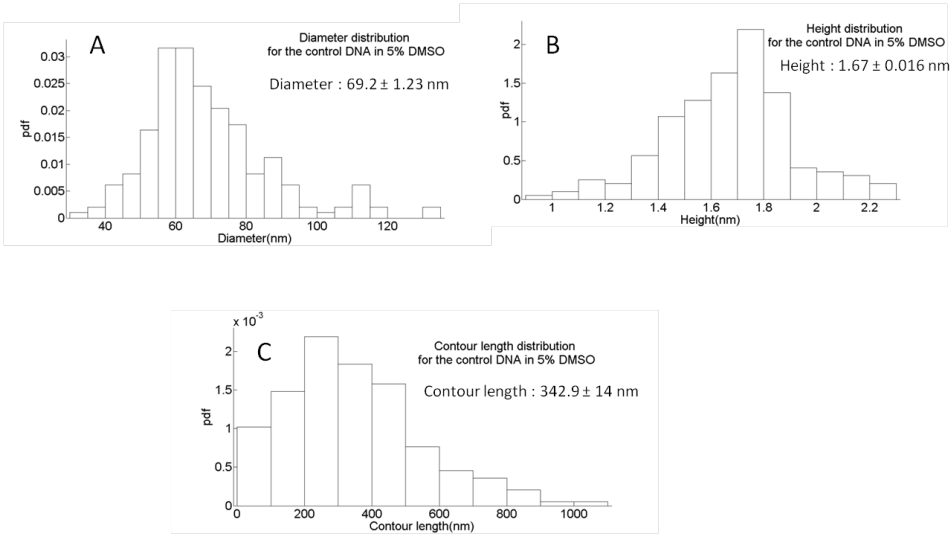


Figure 50. Variation in A. Diameter B. Height C. Contour length of control DNA in 5 % DMSO.

Figure 50 gives the range of variation of contour length, diameter and height of the control DNA in 5% DMSO. The contour length of the control DNA decreased from the average value of 1040 nm to 342.9 nm. The range for this variation varied from around 57nm to 1020 nm.

Also, the height and diameter measurements are very high as compared to the average value of height and diameter of the DNA.

The plots for individual blobs' contour length, diameter and height can give a better idea of what is happening with the molecule.

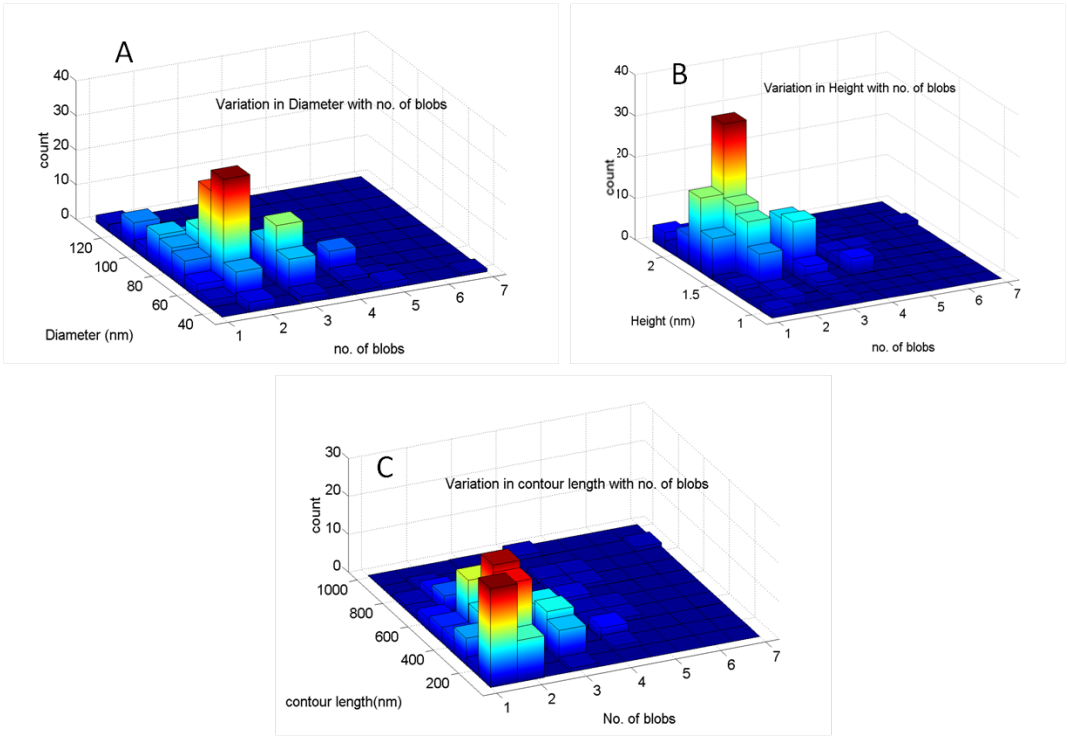
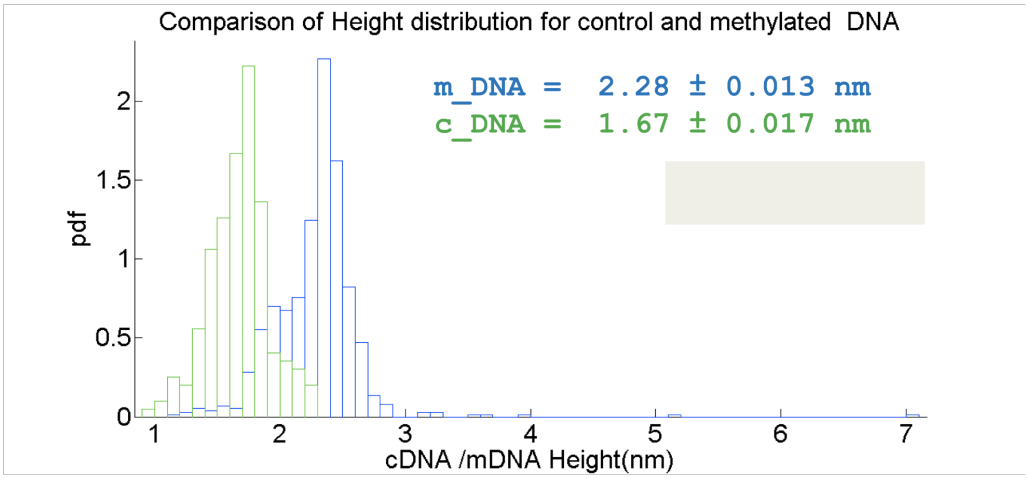


Figure 51. Variation in A. Diameter with the no. of blobs. B. Variation in Height with the no. of blobs. C. Variation in contour length with the no. of blobs.

Its apparent from the 3D plots of variation in height and diameter (Refer Figure 51), that the average height and diameter are the highest for the blob with 2 frequency. From the variation in contour length, it is seen that the single blob has an average of the smallest contour length and with the increase in the number of blobs; the contour length is also increasing. Plotting average contour length for each blob (Refer figure 52.), shows a trend of increase in contour length with the increase in the no. of blobs.

Comparing the height and diameters for both methylated and control DNA (Refer figure 53), it can be seen that both height and diameter of methylated DNA are higher as compared to the control DNA.



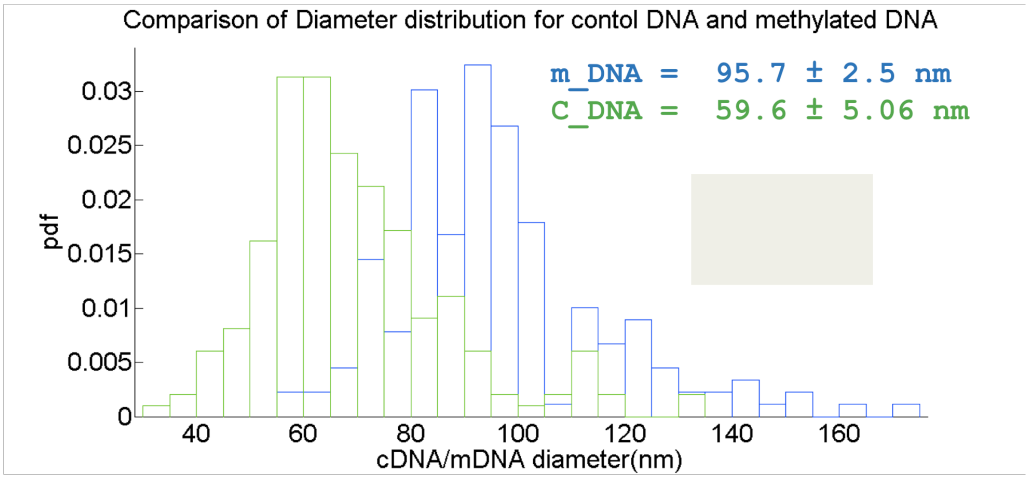


Figure 52 Comparison of Height and Diameter for control and Methylated DNA

From the height and diameter distribution comparison as well as the images (Refer Figure 49, 50), it is very clearly seen that the methylated DNA is becoming more and more condensed structure. But, the question here is that why is the DNA condensing in DMSO and why does methylated DNA condenses more as compared to control DNA?

Even though the literature suggests DNA to be a stable molecule in solution increasing till 25% DMSO in water [125], but still there is some denaturation taking place. Denaturation of the DNA starts at the ends. This completely relates to the observation from the images (Refer Figure 48) that these globules or blobs are always at the ends. DMSO interferes with the hydrogen bonding of the nitrogen bases and breaks down the structure of the DNA.

If single DNA molecules are condensing or forming these blobs, then the average contour length should decrease with the increase in the number of blobs. But, it is not the case.

As seen from the Figure 53, it's very much apparent that the average contour length is increasing with the number of blobs. Contour length for 1 blob can explain two situations. Case1 is the molecule which has fully condensed and second case is the molecule which has just started to condense on one end forming only a single blob. But, from the plots, it's apparent that it signifies the molecule which has already formed a bog blob i.e. case 1. So, instead of the hypothesis that for a single blob, the average contour length will be somewhere near to the average contour length of the DNA template, the result is totally contradictory. If we take a look at the individual distributions, it is apparent that case 1 is much more prominent. Also combined with this fact is that more than 1 molecules are combining together , giving rise to more blobs in higher contour length molecules, such that 3 or 4 molecules are combining and thus , thus they are giving higher contour length as well as higher no. of blobs.

Increased value of Average height and diameter measurements (refer Figure 50, 51, 52) also suggest that these blobs are not due to a single DNA molecule, but the DNA are coming together to form the blobs. Normal DNA template denatures around the ends and thus the ends of the DNA open up. Mg^{++} ions on the surface of the freshly cleaved mica surface serve as an electrostatic bridge between DNA and the –vely charged phosphate backbone. With the dielectric constant of the DMSO being almost half of that of water, the interaction between DNA and Mg^{++} ions is increased as compared to when it was with water as a solvent. Mg^{++} ions attach to the phosphate backbone, effectively metalizing the DNA.

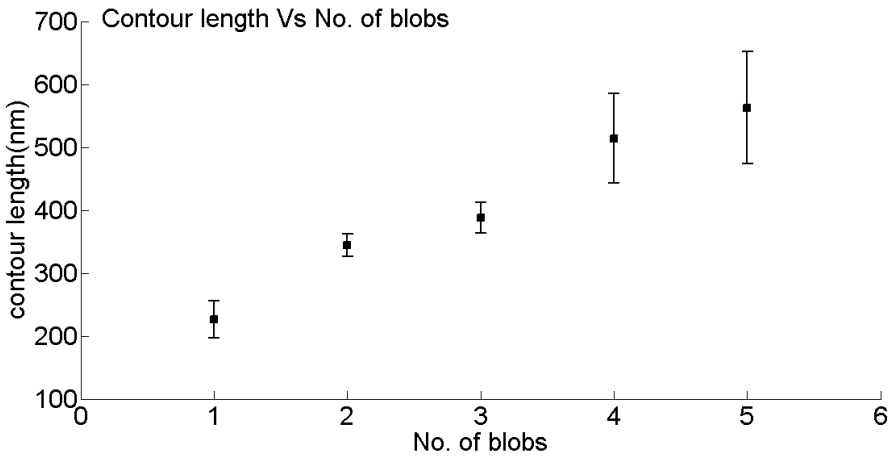


Figure 53. Average Contour length of the DNA molecule with the change in the no. of blobs

Thus, the attached Mg^{++} ions create small globules of –vely charged phosphate backbone around the Mg^{++} ions. As in the data analysis also, it is seen that the blobs first occur at the edges of the DNA polymer. This is explained in Figure54.

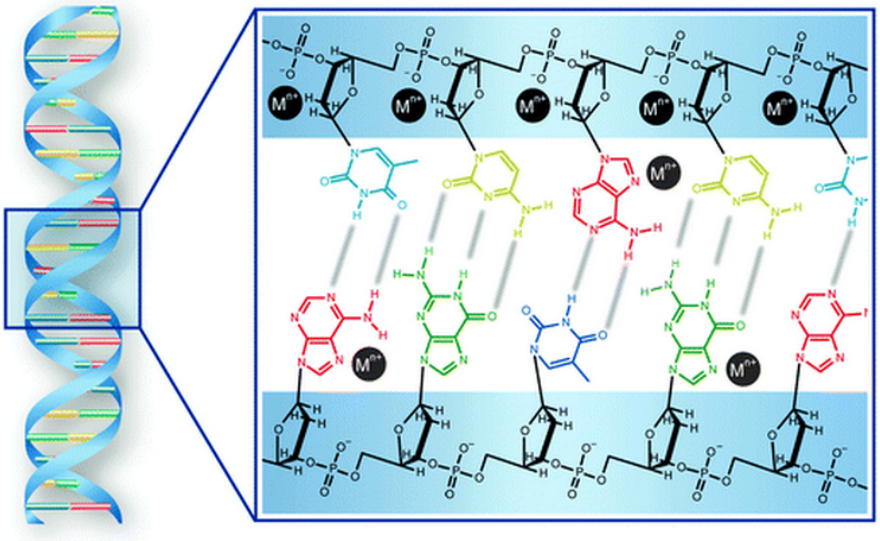


Figure 54. Metallization of DNA

Thus, this clearly explains the blobs formed in the DNA template. This process is very, much a part of the nanowire fabrication and is also termed as DNA metallization. [126][127] But, the case of methylated DNA is a bit different as the images show the DNA folded or condensed into big globular structures. The basic phenomena of the denaturation of DNA around the edges repeat here also, leading to metallization and globules formation. But , the difference in this case is that with the denaturation and opening of the DNA strands, the methyl groups on the cytosine bases get exposed and with this the previous effect gets coupled with the hydrophobic interactions occurring between the methyl groups on the cytosine with those on the DMSO. In simple words, hydrophobic exclusions in which water is excluded and the methyl groups interact among each other, combine and form a larger hydrophobic region. Thus, for methylated DNA, two phenomena couple together and give us huge globular molecule leading to a complete collapse of the molecule. Although, we cannot get any information on the change in the persistence length, due to the collapse of the structure of the DNA, but this experiment certainly gives us a information on how the normal DNA and methylated DNA behaves in different environments.

2.7.3 FUTURE WORK

The aim of this experiment was to observe how DNA and methylated DNA react in a hydrophobic environment and how its persistence length and contour length change with it. Although, we were not able to see the effects due to the breakdown of the structure, but we were still able to get critical information on the different behaviors of both m-DNA and normal DNA in DMSO. Future work in this field will include a hydrophobic solvent in which DNA is completely stable and observe if there are any changes in the interfacial interactions between the Mg-mica surface and DNA.

Chapter 3

STUDY OF THE PHYSICAL DIFFERENCES IN METHYLATED AND NORMAL CHROMATIN AND THEIR INTERACTIONS WITH METHYL CPG BINDING PROTEINS

Project 2

3.1 INTRODUCTION

CpG methylation is an essential requirement for the normal development of mammals, but aberrant changes in the methylation can lead to tumor progression and cancer. An in-depth understanding of this phenomenon can provide insights into the mechanism of gene repression. There are many postulated mechanisms by which DNA methylation represses transcription. Gene transcription in mammals does not involve naked DNA, but occurs on chromatin. Thus, modifications to the chromatin in itself, leads to a great impact on the gene expression. The basic idea is that DNA methylation involves the generation of a chromatin structure that limits access to the region of the genome. There are many possibilities for this to happen. First is by the binding of the methylated region of the chromatin to the methyl CpG binding protein, which recruits HDAC inhibitors and sin3A (transcription repressor) and blocks the regions of the methylated genome from access to the actuators [128], [129].Methyl CpG binding protein is a nuclear protein known for its specific interaction with the methylated DNA, via its methyl binding domain (MBD), leading to gene silencing [129]. This can be explained as in Figure (55).

Another hypothesis for explaining gene silencing by methylation suggests that the methylation can involve the changes in the chromatin structure, leading to a more closed and rigid one. Thus, changes in the DNA around the histone octamer can lead to the chromatin structure getting more compact and thus blocking any access to the methylated region of the genome and blocking transcription .This was already demonstrated by Choy et al [130], using FRET studies on the methylated mononucleosomes.

In order to understand the role of the structural changes and the methyl CpG binding proteins in the chromatin associated with DNA methylation, we carried out studies on the artificially reconstituted chromatin arrays using AFM and recognition imaging.

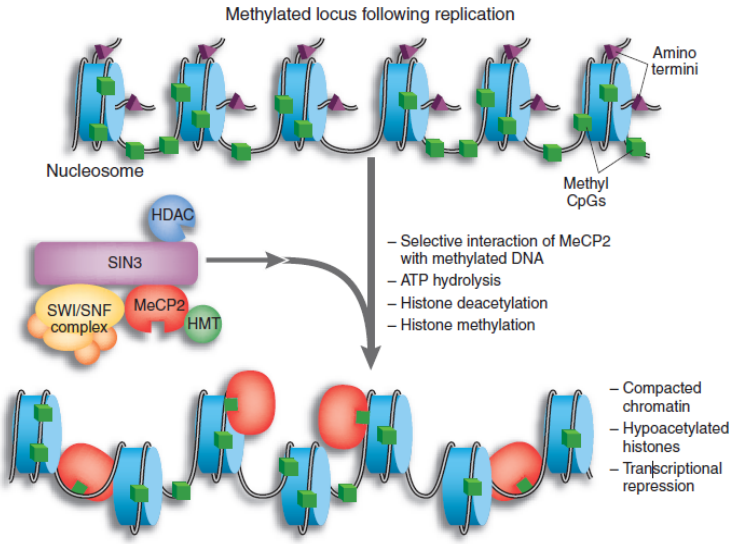


Figure 55. Interaction of meCp2 with the methylated region of the genome. [128]

3.1 MATERIALS AND METHODS

3.1.1 NUCLEOSOME RECONSTITUTION

Htert -601 DNA template used in project 1 was used for this experiment also. Extraction of the DNA template is described in chapter 2. Step salt dialysis method was used for the nucleosomal array reconstitution. [131]. Histone octamer extracted from chicken erythrocytes was used with the DNA template for the artificial reconstitution of the nucleosomal arrays. The DNA and histone octamer were incubated on ice for 30 – 45 min at a ratio of 1 to 1.5 (w/w) in 1xTE buffer (10mM Tris at pH 7.5, 1mM EDTA) to reach a final DNA concentration of 100ng/ul, DTT at 1mM and NaCl at 1M. This solution was further subjected to step wise dialysis into 1M, 0.8M, 0.6M and 0.15 M NaCl solution in 1X TE buffer. The final reconstituted sample after 0.15M NaCl dialysis was dialyzed against 0.25mM TE (pH 7.5) buffer for overnight. All these steps were

carried at 4^0C. Isolated reconstituted sample was crosslinked with 0.1% Glutaraldehyde for imaging in AFM.

3.2.2 ATOMIC FORCE MICROSCOPY

Mica was modified with APTES as described in chapter1. Briefly, mica was freshly cleaved and placed in a desiccator and purged with Argon. 30ul of APTES and 10ul of N, N diisopropylethylamine were added to the centrifuge caps in the bottom of the desiccator. The reaction was carried for 40- 60 mins. Cross linked reconstituted arrays were added to the APTES mica surface and incubated for 40- 60 min and was then rinsed with distilled water gently and dried with nitrogen. The imaging was done with a AFM 5500 (N9410S) with Si_3N_4 cantilevers from AppNano with a spring constant ranging from 25-75 N/m and resonance frequency around 300kHz.

3.3 RESULTS AND DISCUSSION

3.3.1 HTERT- 601 DNA TEMPLATE RECONSTITUTED WITH THE HISTONE OCTAMER EXTRACTED FROM CHICKEN BLOOD CELLS.

The htert promoter- 601 sequence was obtained by digesting the plasmid with Ecor1 and BamH1 followed by gel purification. The samples were reconstituted with the chicken histone octamer and the DNA template, which has htert promoter and 9 repeats of 601 nucleosome positioning sequence [132], [133]. Since every 177 bp are needed for a single nucleosome, the maximum loading number used for the reconstitution was 11. The samples were diluted to 1ng/ul in HEPES buffer, followed by cross-linking with 0.1% glutaraldehyde for 30 mins. Cross linked samples were further immobilized on freshly prepared APTES mica surface. The AFM images of the artificially reconstituted chromatin arrays are shown in Figure56.

Imaging analyses have been performed on this data set. The images of the chromatin show a high loading on one end as compared to the other end, as can be seen in one of the blown up images in Figure 57. Average loading number for the full nucleosomal array is around 10. This is due to the fact that the 601 sequence is the nucleosome positing sequence and htert promoter is

not. Htert Promoter sequence is rich in CpG, but does not load nucleosome that efficiently. Thus, on seeing the chromatin image itself, one can identify the methylation rich and not so rich region. (As shown in the Figure 56)

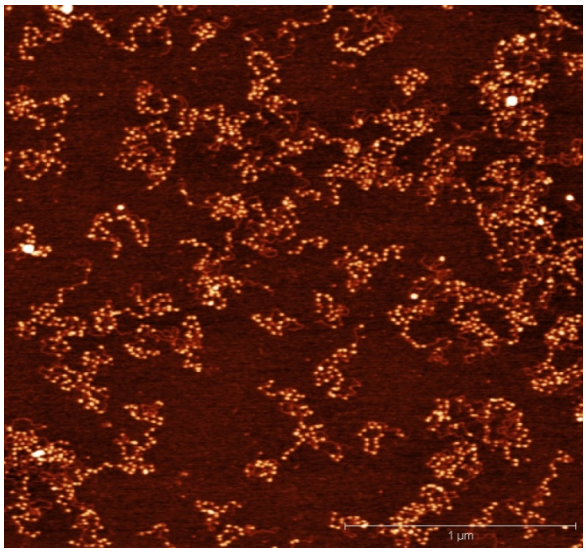
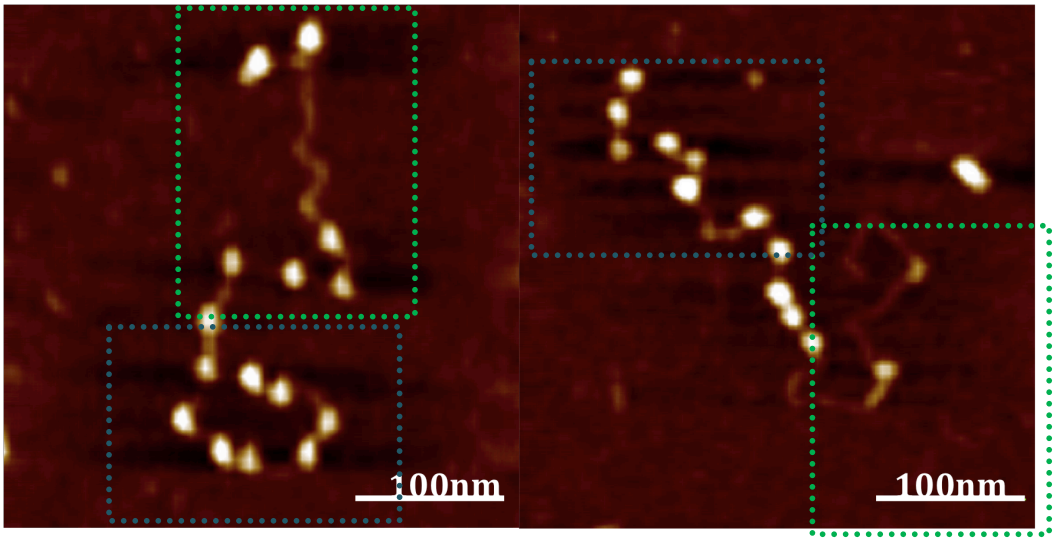


Figure 56. AFM Image of a Reconstituted Nucleosomal Array



Htert promoter ……
601 sequence ……

Figure 57. Topographic blow up of the AFM Image showing htert and 601 sequence.

Further analysis on both the methylated and control reconstituted chromatin was done. Figure 58, 59 show images of the control and methylated chromatin.

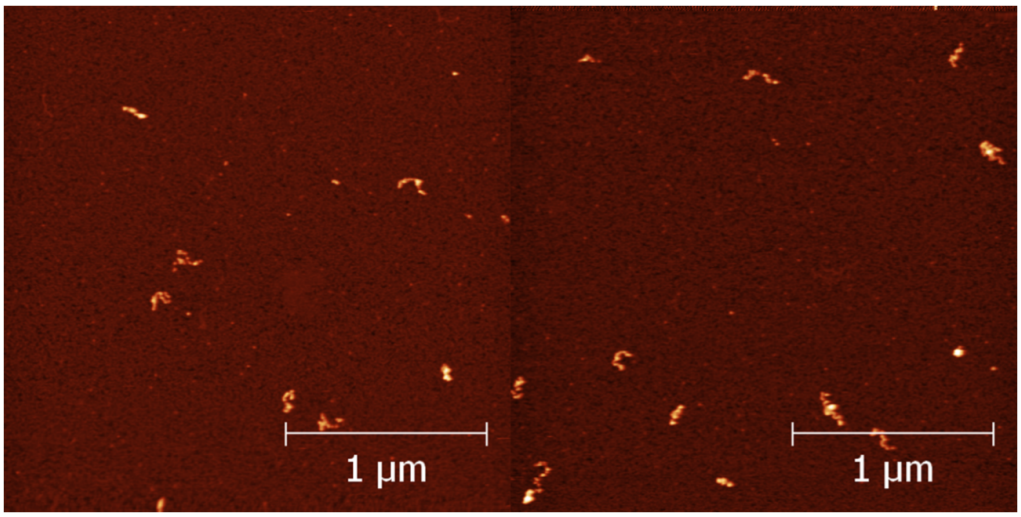


Figure 58. Topographic images of the Methylated Reconstituted Chromatin

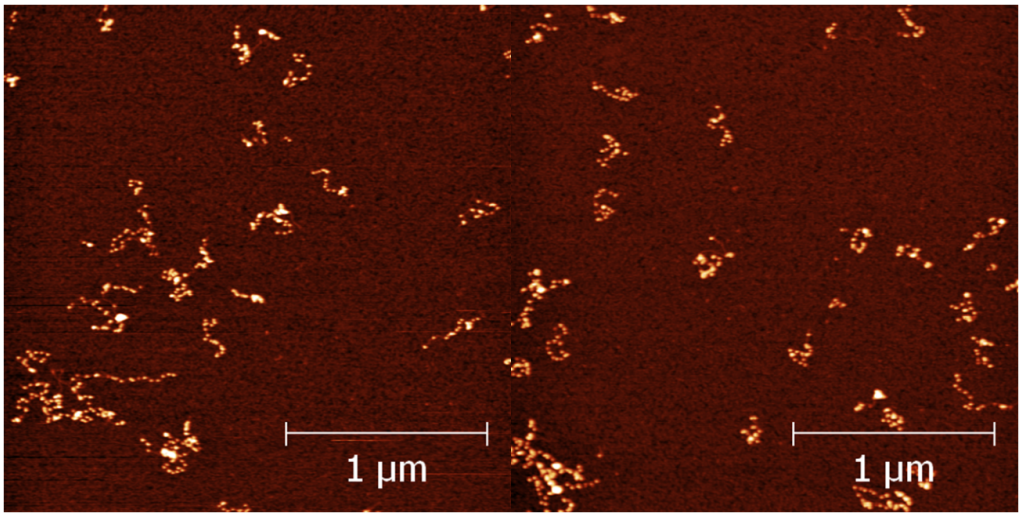


Figure 59. Topographic images of Control Reconstituted Chromatin

As seen from the topographic images of the chromatin samples, the images of the methylated chromatin give an idea that the DNA is more tightly packed around the histones in methylated arrays as compared to the unmethylated ones. It's actually difficult to see individual nucleosomes. To analyze this observation, distribution of the entry and exit angle for both the methylated and control chromatin was studied and is given in Figure 60. The histogram shows a very broad distribution and does not give any significant differences between the two sample sets.

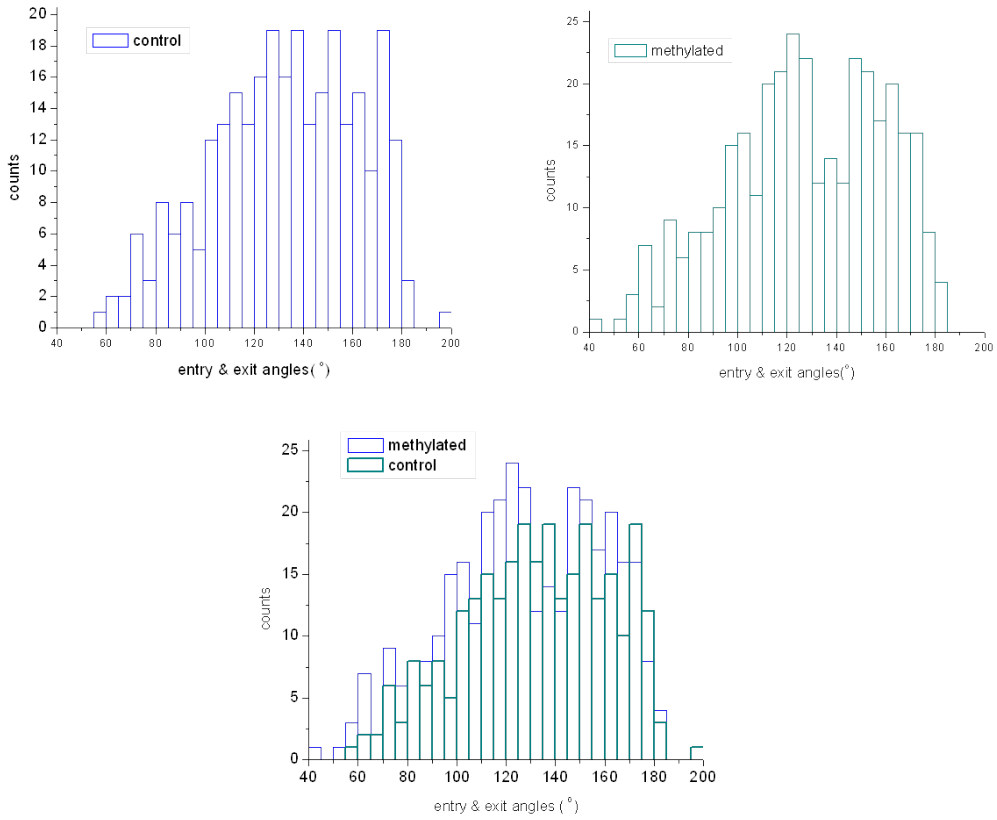


Figure 60. Entry and exit angle histograms for Control and Methylated chromatin

Height and diameter distribution of these images were also done on these images and the methylated chromatin gives different height distribution as compared to the unmethylated chromatin.

Histograms for the height distribution show very distinctly that height of the methylated chromatin is very different than the control chromatin. We are using the same histones and DNA template for the experiment. Also, the height of methylated DNA and the control DNA are almost the same. So, this difference in height is a bit unaccountable. A closer look at the methylated chromatin arrays shows that these are very tightly packed as compared to the extended control chromatin arrays. It looks like they are somehow overlapping on each other as it is a bit difficult to find individual nucleosomes in the arrays. Data for the distribution of the height also relates to this.

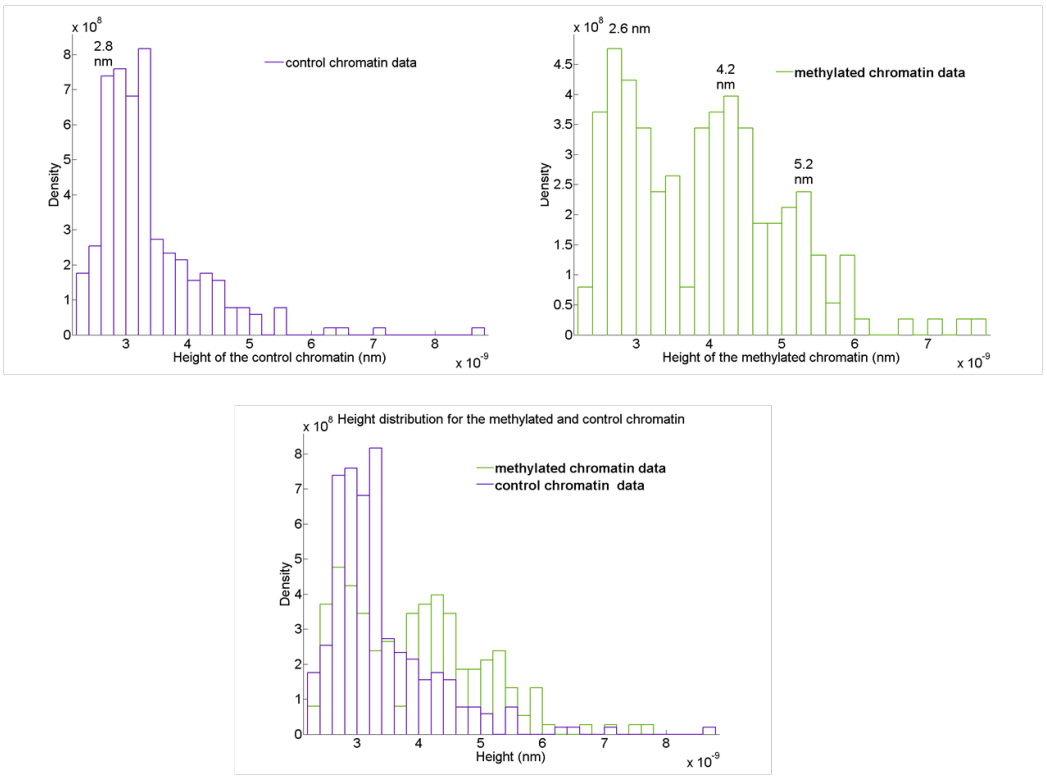


Figure 61.Height Distribution for the Control and Methylated chromatin

As we see, the height of the methylated chromatin has almost three peaks at 2.6 nm, 4.2 nm and 5.2 nm as compared to only one for control chromatin at 2.8 nm. It can be considered that the nucleosomes (histones with the linker DNA) are trying to overlap but are not fully overlapping on each other at a height of 4.2 nm. Heights of 5.2 nm can be explained by observing that it's almost the double height of the single nucleosomes and thus two methylated individual nucleosomes are overlapping on each other.

AFM image analysis of the methylated and unmethylated nucleosomal arrays show that both the arrays have the same loading behavior and no difference in the way the linker DNA enters and leaves the nucleosome. Although, no differences are seen in the entry and exit of the linker DNA in the nucleosome ,agreeing with karymov's data[134] ,but periodic height differences are seen in methylated chromatin as compared to the control chromatin showing an overlap of the nucleosomes one each other. This can also be related to condensing of the structure of the methylated chromatin and thus agree fully with the Choy et al observations. [130]. With the result in

chapter 2 of persistence length of the methylated DNA being higher as compared to the control DNA, making it much more stiffer on the interface of the Mg-mica surface, we can relate this to the current result of the overlapping and compaction of the nucleosomal arrays. With water being more hydrophilic as compared to the Mg-mica surface, the methylated DNA is pushed onto the mica surface. The same correspondence can be used here also. The methylated DNA wrapped around the histone octamer, is pushed onto the histones around which it is wrapped and thus is condensing the chromatin and making it much more rigid. Thus, the results from the persistence length can very well explain the phenomena of chromatin compaction in the case of methylated DNA and we are able to observe that DNA methylation leads to chromatin compaction and thus, can lead to silencing of the genes. Thus, we were able to see differences in the organization and behavior of the methylated chromatin and control chromatin at the nm scale levels.

3.4 RECOGNITION IMAGING OF THE METHYLATED AND UNMETHYLATED CHROMATIN WITH METHYL CPG BINDING PROTEINS

3.4.1 MATERIALS AND METHODS

3.4.1.1 RECONSTITUTION OF THE CHROMATIN SAMPLES

Htert -601 DNA template used in project1 was used for this experiment also. Extraction of the DNA template is described in chapter 2. Step salt dialysis method was used for the nucleosomal array reconstitution. . Histone octamer extracted from chicken erythrocytes was used with the DNA template for the artificial reconstitution of the nucleosomal arrays. The DNA and histone octamer were incubated on ice for 30 – 45 min at a ratio of 1 to 1.5 (w/w) in 1xTE buffer (10mM Tris at pH 7.5, 1mM EDTA) to reach a final DNA concentration of 100ng/ul, DTT at 1mM and NaCl at 1M. This solution was further subjected to step wise dialysis into 1M, 0.8M, 0.6M and 0.15 M NaCl solution in 1X TE buffer. The final reconstituted sample after 0.15M NaCl dialysis was dialyzed against 0.25mM TE (pH 7.5) buffer for overnight. All these steps were carried at 4^0C. Isolated reconstituted sample was cross linked with 0.1% Glutaraldehyde for imaging in AFM.

3.4.1.2 ATTACHMENT OF THE RECONSTITUTED ARRAY ON APTES MICA SURFACE

APTES (3-Aminopropyltriethoxysilane) glutaraldehyde-functionalized mica surfaces were freshly prepared as described earlier [56]. Fifteen micro liters of cross linked nucleosomal arrays (diluted to 0.5 ng/mL in 0.25mM EDTA buffer) were pipetted onto the APTES-GD mica surface. The surface was allowed to stand for 50 min and was later rinsed with EDTA buffer and loaded into the AFM liquid cell for imaging under MAC mode.

3.4.1.3 FUNCTIONALIZATION OF THE PROBES

The AFM probe was functionalized with recombinant methyl CpG binding protein. Cantilevers were ultraviolet-cleaned using an ultraviolet ozone cleaner for 15 min and coated for Mac Mode operation. The cantilevers were introduced into a desiccator containing a 30 ul of freshly distilled APTES (Sigma-Aldrich, St. Louis, MO) in a water-free argon environment, and left for 50 min. One milligram of MAL-PEG 24-NHS ester (MW - 1395, fully stretched length - 9.52 nm; Quanta BioDesign, Powell, OH) was dissolved in 500 ul of chloroform to with 5ul of triethylamine. The APTES-treated cantilevers were rinsed with chloroform and then dipped into the solution containing the linker for 2–3 h.

The meCp2 recombinant protein was thiolated as follows:

A solution of 1 mg/mL of meCp2 was washed in buffer A (100 mM NaCl, 50 mM NaH_2PO_4, 1 mM EDTA, pH 7.5) on a PD10 column and a 10-fold molar excess of SATP in DMSO added to the fractions containing meCp2 protein. The samples were left for 1 h under argon at room temperature, and the samples were again washed with buffer A on a PD10 column. Probes functionalized with MAL-PEG-NHS were washed with chloroform and were dried with argon, and drops of 50 ul SATP-treated antibody solution with 25 ul NH_2OH in 50 mL buffer A (150 mM NaCl, 50mM Na_2HPO_4, pH 7.5) were placed onto each probe and left for 1 h. The tips were washed with buffer A, and then PBS buffer and were stored in PBS buffer at 4^oC. The probe and sample were mounted on a Pico SPM I scanning probe microscope (Agilent Technologies, Chandler, AZ), and topographic and recognition imaging were carried out under EDTA buffer

solution. We used silicon nitride probes with nominal spring constant 0.1 N/m (Veeco Instruments, Plainview, NY) and nominal spring constant 0.08 N/m (Olympus, Melville, NY). Individual probes were calibrated using the thermal noise method. Recognition images were usually acquired at a peak-to-peak amplitude of 8.3 nm and a scan rate of 1.5 lines/s using PicoTREC electronics (Agilent).

3.4.2 RESULTS AND DISCUSSION

Recognition imaging for the artificially reconstituted nucleosomal arrays used in section 3.2.1, was done with methyl CpG binding protein (meCp2) on the functionalized tips. The set up for the experiment is as in figure 62.

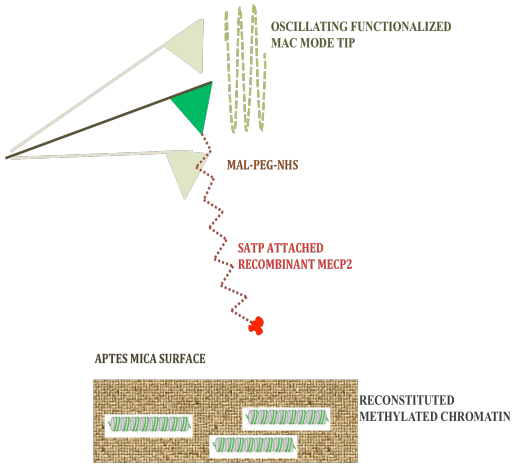


Figure 62. Experimental setup for recognition imaging of the artificial nucleosomal arrays with meCp2 protein.

Figure 63 gives the topographic and the recognition images (simultaneously acquired) of the methylated nucleosomal arrays taken with a MAC lever functionalized with meCp2 protein.

Topographical images (left) show individual nucleosomes appearing as white spot on a dark background and the recognition images (right), signifying the interaction between the meCp2 protein and the methylated nucleosomal array, i.e., individual recognition events as dark spots on a lighter background.

TOPOGRAPHIC IMAGE OF METHYLATED CHROMATIN

RECOGNITION IMAGE OF METHYLATED CHROMATIN

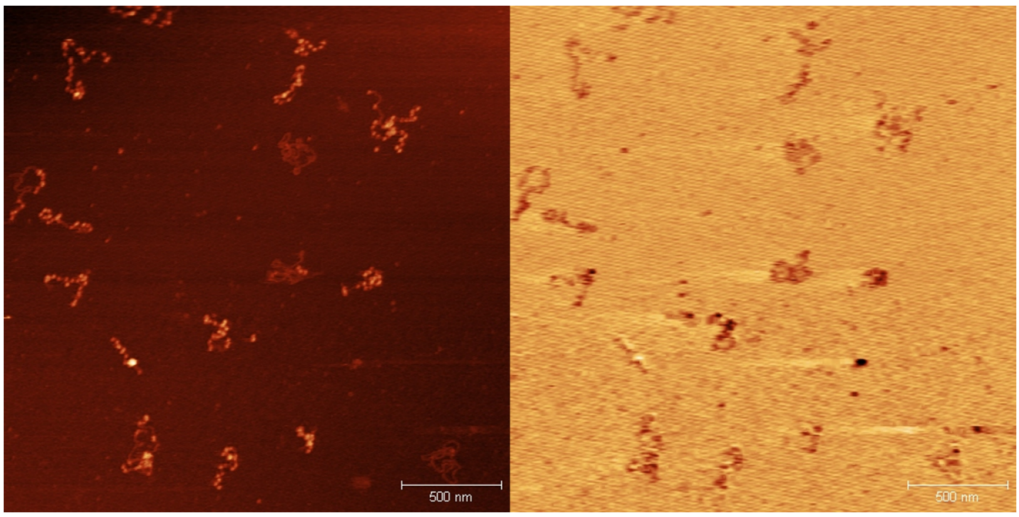


Figure 63. Topographic and Recognition data for the methylated reconstituted nucleosomal array with SATP modified meCp2 recombinant protein

Analysis of both topographic and recognition images demonstrate that the recognition events coincide with the location of nucleosomes. Detailed analysis of the recognition imaging shows a good recognition efficiency. After blocking the meCp2 on the tip with addition of methylated DNA in solution, the recognition signal was abolished, indicating the interaction to be specific (Figure64).

The sequence of the DNA template was selected in such a way that it was CpG rich on one end as compared to the other end (Sequence –htert promoter- 601 sequences). Referring to the sequence, it was expected that we will have more recognition towards one end, but from the recognition image in figure 63, the recognition looks uniform all over. To make sure that, this is the real case scenario, dips in amplitudes were measured for few individual nucleosomes in their recognition images.

TOPOGRAPHIC IMAGE OF METHYLATED CHROMATIN

RECOGNITION IMAGE OF METHYLATED CHROMATIN

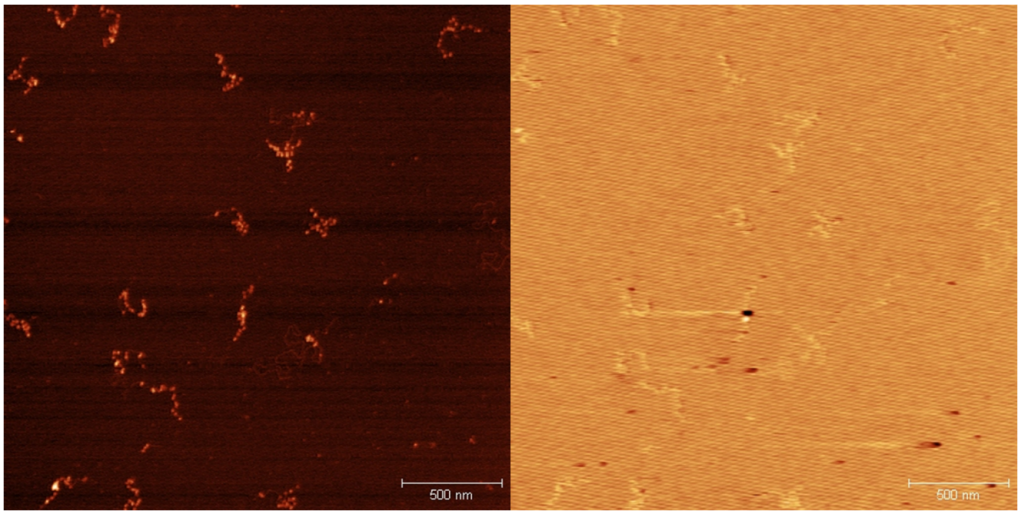


Figure 64. Specificity control for methylated chromatin with meCp2 functionalized tips

Plots in figure 65, show that there is more dip in amplitude towards the htert region of the chromatin as compared to the 601 region. Thus, establishing the fact that meCp2 is interacting strongly with the more methylated region of the DNA, i.e., htert promoter as compared to the one with lesser methylated region, 601 repeat sequences.

Recognition imaging was also done on the control chromatin with the same meCp2 functionalized tip. Figure 65 shows topographic and recognition images of the samples. As seen from the images, meCp2 does not recognize the control chromatin. No recognition signal was seen for these reconstituted arrays. Thus, establishing that meCp2 specifically binds with the methylated region of the genome.

3.4.3 FUTURE WORK

The conformity of the phenomena of the interaction of the meCp2 with methylated DNA in reconstituted artificial arrays and compaction of these arrays in methylated samples is just a starting step towards the study of interactions of meCp2, methylated DNA and chromatin in gene silencing and cancer progression. This opens up future pathways in understanding these at nm

scale. It will be much more interesting to study these interactions in real chromatin extracted from cell lines.

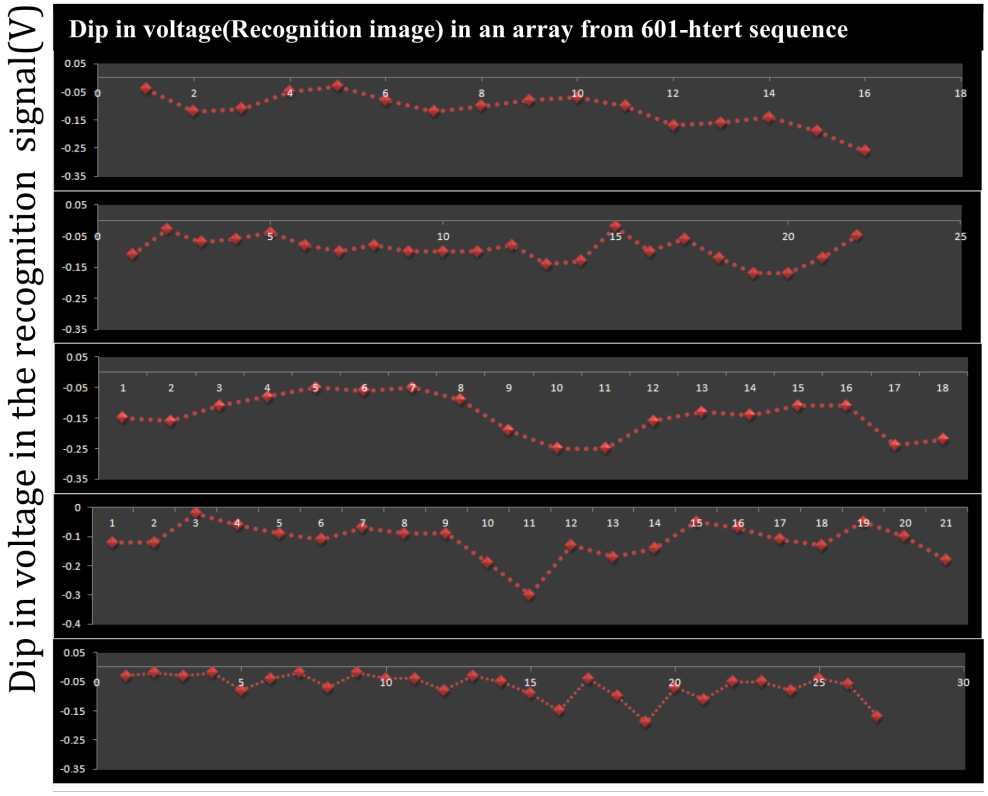


Figure 65. Dip in amplitude in recognition image of the individual arrays from Figure 62.

Establishment of the topographical differences in the methylated chromatin and normal chromatin can help in identifying these in the euchromatin or heterochromatin by only taking their topographical images using AFM and also explains the phenomena of chromatin compaction in DNA methylation. It will be all the more interesting and intriguing to expand this study from normal cell lines to cancer cell lines.One of the major causes of cancer progression and tumor growth is the aberrant methylation patterns observed in the genome. With the meCp2 recognizing the methylated region of the genome, the methylation patterns can be observed from the recognition maps of the chromatin extracted from the cell lines, which can give very critical information on the gene silencing. Also, it can be observed how different reagents change the

methylation patterns in the genome and thus can be a helping step towards the ongoing research on the drug discovery for cancer and tumors progression.

TOPOGRAPHIC IMAGE OF CONTROL CHROMATIN

RECOGNITION IMAGE OF CONTROL CHROMATIN

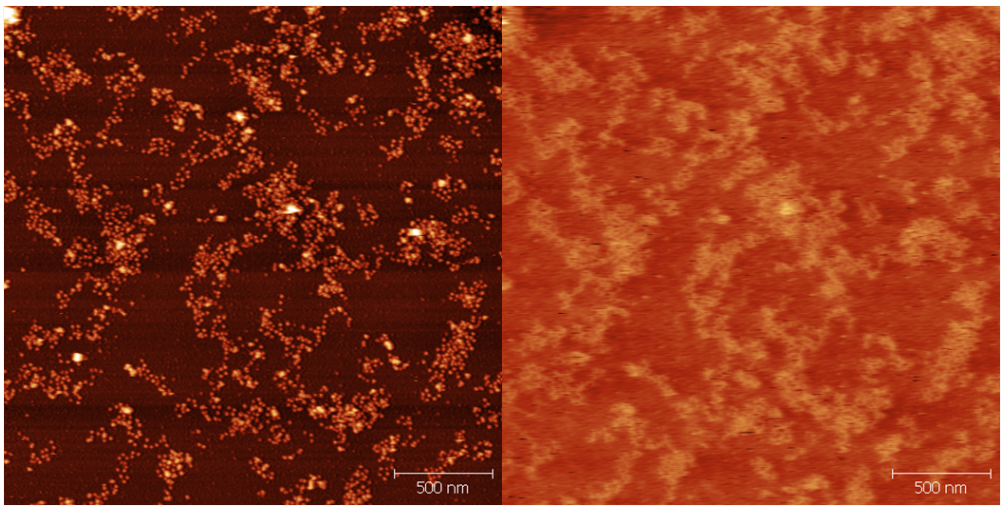


Figure 66. No Recognition Signal with the AFM Tip Functionalized With meCp2 on the Control Chromatin

3.5 IDENTIFICATION OF METHYL CPG BINDING PROTEINS (MECP2) IN THE CHROMATIN EXTRACTED FROM RAT BRAIN CELLS.

3.5.1 MATERIALS AND METHODS

3.5.1.1 ISOLATION OF THE CHROMATIN FROM RAT BRAIN CELLS

Mononucleosomes were extracted from the rat brain cells by the action of Micrococcal nuclease digestion. (From Juan Ausio lab) Isolated samples were cross linked with 0.1% Glutaraldehyde for imaging in AFM.

3.5.1.2 ATTACHMENT OF THE SAMPLES ON APTES MICA SURFACE

APTES (3-Aminopropyltriethoxysilane) glutaraldehyde-functionalized mica surfaces were freshly prepared as described earlier (13). Fifteen micro liters of cross linked nucleosomal arrays (diluted to 0.5 ng/mL in HEPES buffer) were pipetted onto the APTES-GD mica surface.

The surface was allowed to stand for 50 min and was later rinsed with HEPES buffer and loaded into the AFM liquid cell for imaging.

3.5.1.3 TOPOGRAPHIC IMAGING IN AIR

Topographic imaging was done on the APTES micas surface using AFM 5500 from Agilent Technologies using Si_3N_4 cantilevers with spring constant ranging from 25- 75N/m and resonating around 300 kHz.

3.5.1.4 FUNCTIONALIZATION OF THE PROBES AND RECOGNITION IMAGING

The AFM probe was functionalized with anti-mecp2 (sigma). Cantilevers were ultraviolet-cleaned using an ultraviolet ozone cleaner for 15 min and coated for Mac Mode operation. The cantilevers were introduced into a desiccator containing a 30 ul of freshly distilled APTES (Sigma-Aldrich, St. Louis, MO) in a water-free argon environment, and left for 50 min. One milligram of MAL-PEG 24-NHS ester (MW - 1395, fully stretched length - 9.52 nm; Quanta BioDesign, Powell, OH) was dissolved in 500 ul of chloroform to with 5ul of triethylamine. The APTES-treated cantilevers were rinsed with chloroform and then dipped into the solution containing the linker for 2–3 h.The anti-meCp2 was thiolated as described in the previous section. Probes functionalized with MAL-PEG-NHS were washed with chloroform and were dried with argon, and drops of 50 ul SATP-treated antibody solution with 25 ul NH_2OH in 50 mL buffer A (150 mM NaCl, 50mM Na_2HPO_4, pH 7.5) were placed onto each probe and left for 1 h. The tips were washed with buffer A, and then PBS buffer and were stored in PBS buffer at 4^oC. The probe and sample were mounted on a Pico SPM I scanning probe microscope (Agilent Technologies, Chandler, AZ), and topographic and recognition imaging were carried out under HEPES buffer solution. We used silicon nitride probes with nominal spring constant 0.1 N/m (Veeco Instruments, Plainview, NY) and nominal spring constant 0.08 N/m (Olympus, Melville, NY). Individual probes were calibrated using the thermal noise method. Recognition images were usually acquired at a peak-to-peak amplitude of 8.3 nm and a scan rate of 1.5 lines/s using PicoTREC electronics (Agilent).

3.5.2 RESULTS AND DISCUSSION

Isolated mononucleosomes were obtained from the chromatin extracted from rat brain cells. The samples were diluted to 1ng/ul in HEPES buffer, followed by cross-linking with 0.1% glutaraldehyde for 30 mins. Cross linked samples were further immobilized on freshly prepared APTES mica surface. The AFM topographic images of the mononucleosomes obtained on the freshly prepared mica surface are as shown in figure (65)A and B. Figure 65A gives the image of mononucleosomes that are 25% enriched in meCp2 whereas Figure 65B gives the image of the control sample of mononucleosomes without any meCp2 enrichment.

A B

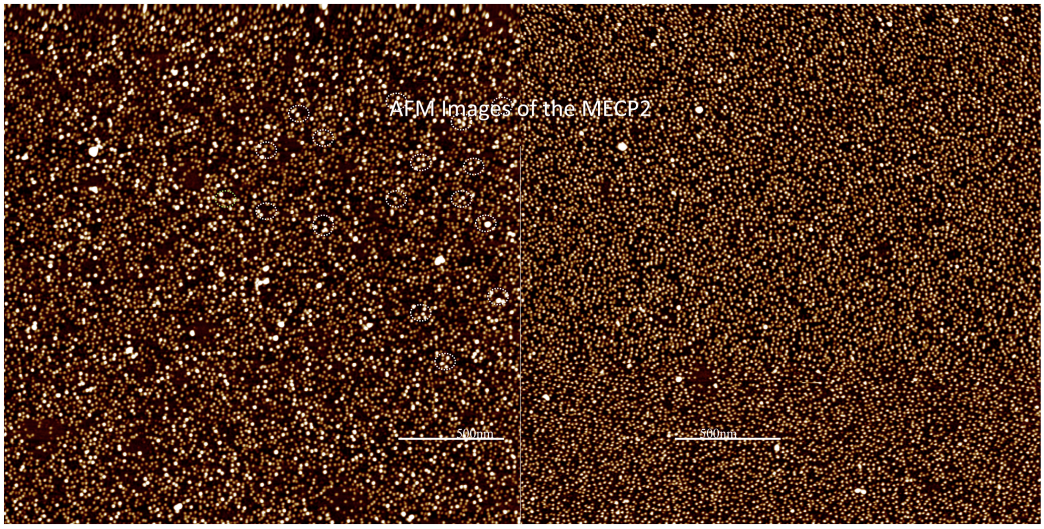


Figure 67. Topographic images of the mononucleosomes extracted from rat brain cells. A. 25% mecp2 enriched. B. Control sample.

Image analysis of figure 65 A shows a higher density of brighter mononucleosomes. Brighter mononucleosomes correspond to a higher height as compared to the lesser brighter ones. Figure 65B shows a uniform population of mononucleosomes wrt height. Height distribution of these samples is plotted in Figure 66.

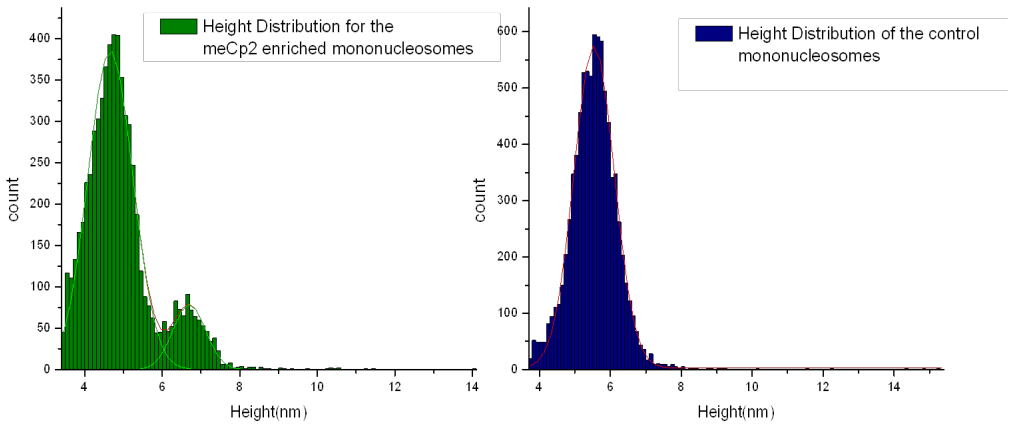


Figure 68. Height Distribution for the meCp2 enriched and control mononucleosomes.

From the data analysis for the height distribution, it is very distinctly seen that meCp2 enriched mononucleosomes have a mixed population of two heights with a height distribution of:

Height 1: 4.6 ± 1.1 nm

Height 2: 6.7 ± 0.8 nm

The second sample, control, has a height distribution of:

Height: 5.54 ± 1.1 nm

Also from the figures, it is seen that the population of the samples with higher height for meCp2 enriched mononucleosomes is almost 25 % of the total population. This is exactly the % of meCp2 enriched mononucleosomes. Thus, we were able to identify the meCp2 enriched mononucleosomes in the chromatin extracts from rat brain cells.

In order to be sure that these are meCp2 enriched, recognition imaging on these samples was done with anti-meCp2 antibody, which recognizes meCp2. Figure 67 shows the experimental setup for the control reaction.

Topographic images are shown in figure with the simultaneously acquired recognition image in figure 68A The dark spots in the recognition image mark regions where the antimeCp2 binds to the mononucleosomes. The recognition efficiency is not that good as expected. If our assumption is true only mononucleosomes with higher heights should be giving recognition signal. As seen from the image the efficiency for recognition is only 12 %.

On measuring the heights of mononucleosomes with recognition signals, it was seen that only mononucleosomes with a higher height above 5 nm gave the recognition signal. All the lower heights mononucleosomes did not give any recognition signal. Thus, anti- meCp2 is only recognizing mononucleosomes with meCp2 attached to them. The lower heights mononucleosomes do not give any recognition signal. This, in itself serves a s a specificity proof of the recognition imaging of the meCp2 enriched mononucleosomes with the anti-meCp2 antibody. Figure 68B shows that only higher heights mononucleosomes are giving the recognition signal. Thus, we were able to successfully identify the meCp2 on the mononucleosomes and perform recognition on them.

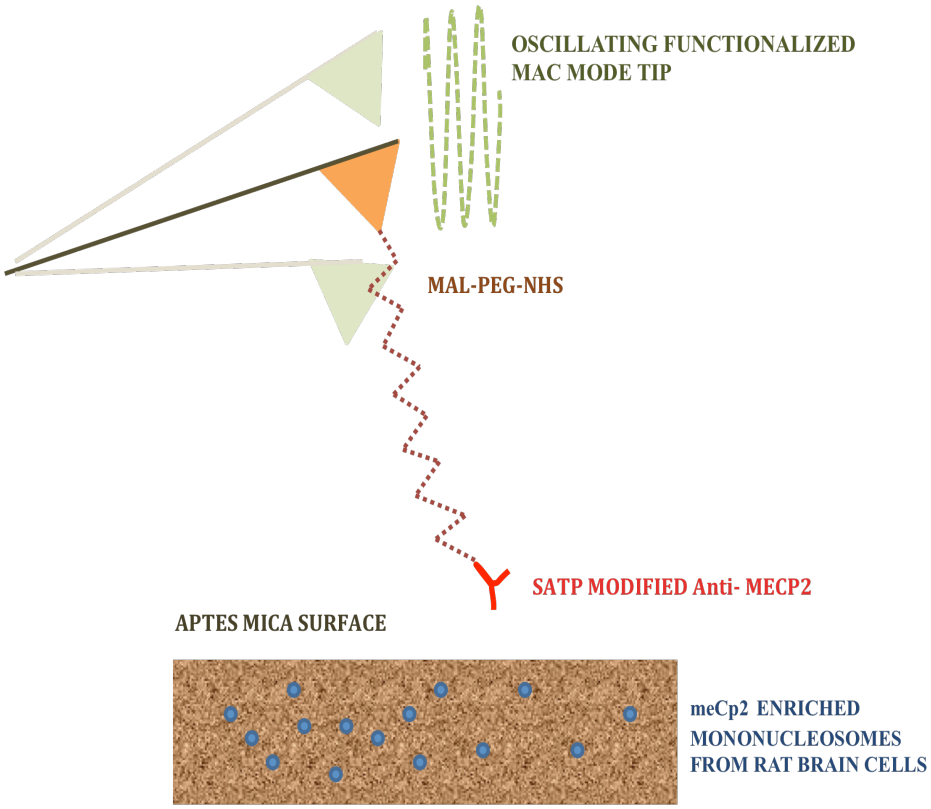
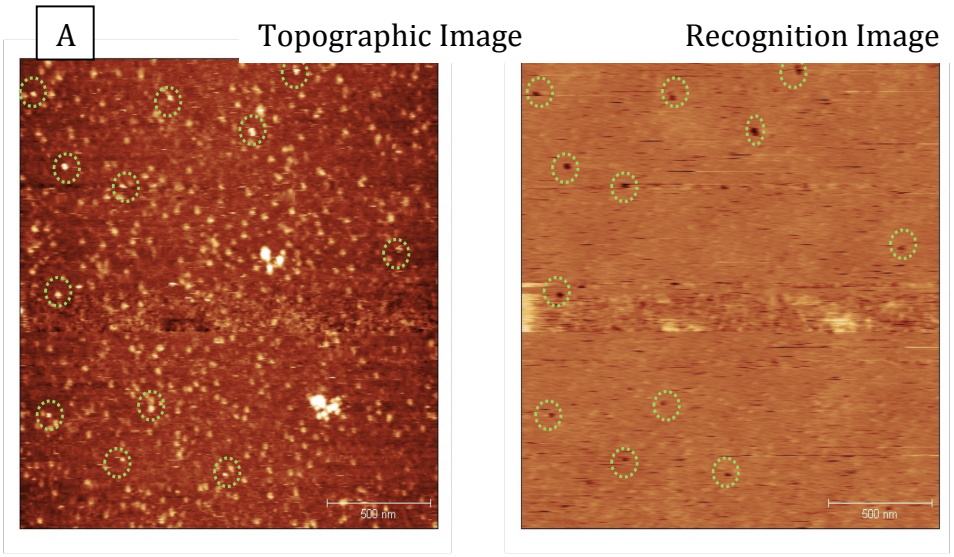


Figure 69. Experimental setup for meCp2 identification by anti-meCp2

Recognition Imaging with Anti-meCp2 on the tip and meCp2 enriched mononucleosomes on the sample surface



Almost 12% recognition signal

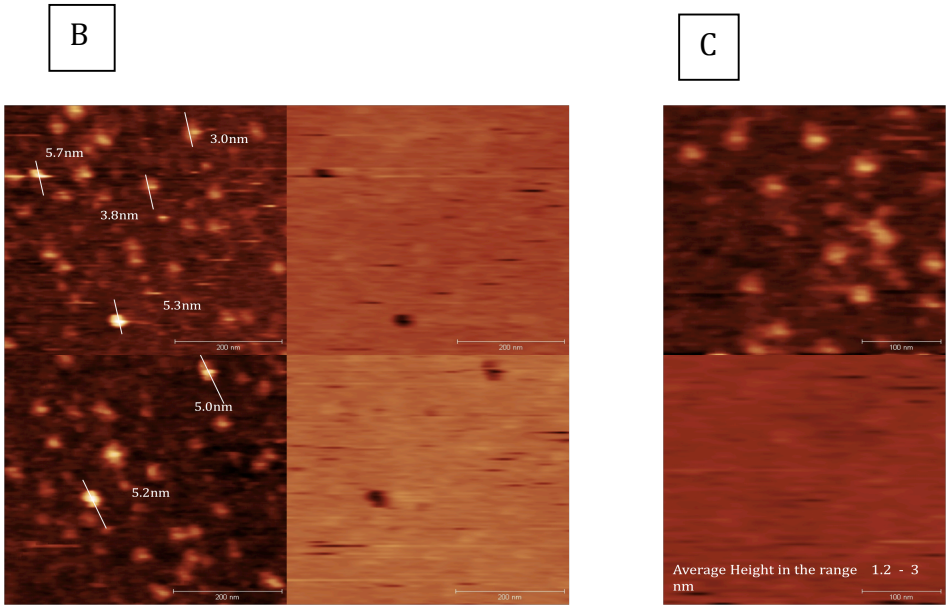


Figure 7065 A. Simultaneously acquired topographic and Recognition images of the meCp2 enriched mononucleosomes. B. Only higher height (> 5 nm) mononucleosomes showing

recognition as compared to the lower heights ones (< 5 nm) C. No recognition for the control mononucleosomes.

3.5.3 FUTURE WORK

With this project we were successfully able to map the meCp2 on the mononucleosomes extracted from the rat brain cells. Utilizing this technique in the near future, we can identify the percent of meCp2 rich regions, which is a marker of DNA methylation, in the chromatin extracted from human cell lines, including cancer cell lines. Direct visualizations of the meCp2 enriched regions of the human genome can be done by AFM. Chromatin extracted from various cell lines can be identified for the enrichment of meCp2 and thus, methylation. This can again give us critical information on the reaction pathways of meCp2 in the live cell processes and its role in cancer progression.

CONCLUSION

With this thesis, we were successfully able to map out differences between methylated DNA and normal DNA and explain the phenomena of silencing of the genes as an interfacial effect. Further, we were able to observe distinct changes in the topography images of the artificially reconstituted chromatin samples. Methylated chromatin arrays showed higher periodic heights and overlap of the individual nucleosomes on each other leading to a compact structure. As noted in the literature, most of the region of the heterochromatin is in condensed form. It is also known that the heterochromatin mainly consists of methylated regions in its genome and thus is able to silence the gens it is encoding for. This can explain how the heterochromatin adopts a folded, condensed structure and can also explain how the genes are silenced leading to cancer progression. Also, we were able to successfully map meCp2 to the methylated chromatin arrays, thus, establishing the fact that meCp2 interacts specifically with the methylated region of the genome. Lastly, the studies were shifted to the chromatin extracts from the rat brain cells and successful mapping of the meCp2 enriched regions was done both on the topographic and recognition levels. With these studies, we have successfully layed down future pathways for understanding the cascade of events leading to gene silencing involving DNA methylation, chromatin and meCp2.

REFERENCES

1. G.K.Binnig, C.F.Quate, and C.Gerber, March 3 1986, Atomic force microscope, Physical Review Letters, vol. 56, pp. 930–933.

2. N.C.Santos and M.A.R.B. Castanho, Feb 2004, An overview of biophysical applications of atomic force microscopy, Biophysical Chemistry, vol.107, pp 133-149.

3. Hansma and Hoh, 1994 ,Biomolecular imaging with the Atomic force Microscope.Annual Review of Biophysics and Biomolecular Structure Vol. 23: 115-140

4. Shao, Mou, Czajkowsky, Yuan, 1996, Biological Atomic force microscopy-what is achieved and what is needed?, Advances in physics, vol. 45, Issue 1, page.1.

5. A.Alessandrini and P. Facci, 2005, AFM: a versatile tool in biophysics. Measurement science & technology, volume 16, Issue 6, p. R65.

6. A. A. G. Requicha 2003. Nanorobots, NEMS, and nanoassembly, Proceedings of the IEEE, vol. 91, no. 11, pp. 1922–1933.

7. M. Sitti, June 2004, Atomic force microscope probe based controlled pushing for nanotribological characterization, IEEE-ASME T. Mech., vol. 9, pp. 343–349.

8. K. Tsukamoto, S. Kuwazaki, K. Yamamoto, T. Ohtani, and S. Sugiyama, March 2006.Dissection and high-yield recovery of nanometre- scale chromosome fragments using an atomic-force microscope, Nanotechnology, vol. 17, pp. 1391–1396.

9. Binning et al, 1982. Scanning Tunneling microscope, United States Patent, Patent No. 4343993

10. J. F. Nierengarten and S. Setayesh, 2006, Towards polymerizable fullerene derivatives to stabilize the initially formed phases in bulk- heterojunction solar cells, New Journal of Chemistry, vol. 30, no. 3, pp. 313–316.

11. E. Granot, F. Patolsky, and I. Willner, May 2004. Electrochemical Assembly of a CdS Semiconductor Nanoparticle Monolayer on Surfaces: Structural Properties and Photoelectrochemical Applications., The journal of physical chemistry. B, volume 108, Issue 19, p. 5875.

12. J. Hahm and S. J. Sibener ,2000,Time-resolved atomic force microscopy imaging studies of asymmetric PS-b-PMMA ultrathin films: Dislocation and disclination transformations,defect mobility, and evolution of nanoscale morphology , The Journal of chemical physics, volume 114, Issue 10, p. 4730.

13. Hobbs, J. K. 2001, In-Situ Atomic Force Microscopy of Polyethylene Crystallization. 1. Crystallization from an Oriented Backbone., Macromolecules, volume 34, Issue 16, p 5508.

14. Zhong, Inniss 1993, Fractured polymer/silica fiber surface studied by tapping mode atomic force microscopy. Surface Science letters, Volume 290, Issues 1-2,pp L688-L692.

15. Hansma, 1994, Tapping mode atomic force microscopy in liquids, Applied physics letters, Volume 64, Issue 13, p.1738.

16. Hinterdorfer, 1996, Detection and localization of individual antibody-antigen recognition events by atomic force microscopy , PNAS: Proceedings of the National Academy of Sciences, Volume 93, Issue 8, p3477.

17. Shaune J. Hanley, Jul ie Giasson, Jean-Franqois Revol and Derek G. Gray, 1991, Atomic force microscopy of cellulose microfibrils: comparison with transmission electron microscopy. Polymer (Guilford), Volume 33, Issue 21, p4639.

18. J.B. Heymann, 1999, Charting the Surfaces of the Purple Membrane., Journal of structural biology, Volume 128, Issue 3, p243.

19. S.M. Lindsay, 2004, Single molecule Recognition imaging microscopy. PNAS: Proceedings of the National Academy of Sciences, 10.1073/pnas.0403538101.volume 101, Issue 34

20. W.T.Johnson, H.Wang, G.Kada, S.M. Lindsay, 2005, Simultaneous Topography and Recognition Mapping with PicoTREC: A powerful new technology that can be used to map nanometer scale molecular binding sites on a variety of surfaces. NSTI- Nanotech 2005, www.nsti.org, ISBN 0-9767985-1-4, Vol. 2, 2005.

21. P.Hinterdorfer and Z.Reich, Molecular recognition force microscopy: from simple bonds to complex energy landscape, in Nanotribology and Nanomechanics, 2nd ed. B.Bhushan, Ed.Heidelberg: Springer-Verlag, pp.279-303, 2008.

22. K ienberger, Gruber, Hinterdorfer, APPLIED SCANNING PROBE METHODS II NanoScience and Technology, 2006, 143-164, Dynamic Force Microscopy and Spectrocopy.

23. Gaub, 1994, Adhesion Forces Between Individual Ligand-Receptor Pair, Science (New York, N.Y.), Volume 264, Issue 5157, p415.

24. Hinterdorfer, P.; Baumgartner, W.; Gruber, H. J.; Schilcher, K.; Schindler, H. 1996, Detection and localization of individual antibody antigen recognition events by atomic force microscopy. Proc.Natl. Acad. Sci. U.S.A., Volume 93, Issue 8, p3477.

25. Evan Evans, 2001, Probing the relation between force, lifetime and chemistry in single molecular bonds. Annual review of biophysics and biomolecular structure, Volume 30, Issue 1, p105.

26. Gaub, 2002, Force Spectroscopy with single biomolecules. Chemphyschem, Volume 3, Issue 3, p255.

27. Abramovitch, Andersson, Pao, Schitter.2007, A tutorial on the Mechanisms, Dynamics and Control of Atomic force Microscopes. , American Control Conference (ACC), p 3488

28. Roduit, C. "AFM figures" 2010, www.freesbi.ch, Creative Commons Attribution

29. Han and Lindsay, 1998, MAC Mode Atomic Force Microscope for Precision Interfacial Force Measurements. Applied physics letters, Volume 72, page 1656.

30. Han and Lindsay, September 1997, Strained DNA is kinked by low concentrations of Zn2+. Proc. Natl. Acad. Sci. USA, Vol. 94, pp. 10565–10570.

31. Han and Lindsay, 23 December 1996, A magnetically driven oscillating probe

32. Revenko and Proksch, 1 January 2000, Magnetic and acoustic tapping mode microscopy of liquid phase phospholipids bilayers and DNA molecules. J. Appl. Phys., Vol. 87, No. 1.

33. Meyer and Amer, 1988, A Novel optical approach to atomic force microscope. Appl Phys Lett 53: 1045-1047.

34. http://www.sciencephoto.com/media/153285/enlarge

35. MarcoTortonese,1997,Cantilevers and tips for atomic force microscope.

36. Kromka and Rezek,2011,Optimizing atomic force microscopy for characterization of diamond-protein interfaces.

37. W. Travis Johnson, Attaching Antibodies to AFM Probes with the Sulfhydryl Reactive PEG Tether, NHS-PEG18-PDP.Application Note:http://cp.literature .agilent.com/litweb/pdf/5989-7702EN.pdf

38. Hinterdorfer, P. et al. 2000. Poly(ethylene glycol):an ideal spacer for molecular recognition force microscopy/spectroscopy. Single Mol. 1: 99–103.

39. Ferry Kienberger, Vassili Ph. Pastushenko, Gerald Kada, Hermann J. Gruber, Christian Riener, Hansgeorg Schindler and Peter Hinterdorfer, 2000: Static and Dynamical Properties of Single Poly(Ethylene Glycol) Molecules Investigated by Force Spectroscopy : Single Mol. 1 (2000) 2, 123-128

40. D.Lohr, R.Bash, H.Wang, J.Yodh, S.Lindsay. March 2007,Using atomic force microscopy to study chromatin structure and nucleosome remodeling. Methods (San Diego, Calif.) Volume 41 , Issue 3, Page 333.

41. Dalal Y, Wang H, Lindsay S, Henikoff S (2007) ,Tetrameric structure of centromeric nucleosomes in interphase Drosophila cells. PLoS Biol 5(8): e218. doi:10.1371/ journal.pbio.0050218

42. Hongda Wang, Yamini Dalal, Steven Henikoff and Stuart Lindsay, 2008, Single-epitope recognition imaging of native chromatin. Epigenetics & Chromatin, Volume 1, Number 1, 10.

43. Liyun Lin, Qiang Fu,Berea A. R. Williams,Abdelhamid M. Azzaz,Michael A. Shogren-Knaak, John C.Chaput, and Stuart Lindsay, September 2009 ,Recognition Imaging of Acetylated Chromatin Using a DNA Aptamer: Biophysical Journal Volume 97 1804–1807

44. Hinterdorfer, P.; Schilcher, K.; Baumgartner, W.; Gruber, H. J.; Schindler, H. A mechanistic study of the dissociation of individual antibody-antigen pairs by atomic force microscopy. Nanobiology 1998, 4, 39-50.

45. Grubmuller,H., Heymann,B., Tavan,P .Ligandbinding: Molecular mechanics calculation of the streptavidin-biotin rupture force. Science 1996, 271, 997-999.

46. Willemsen, O. H.; Snel, M. M. E.; van der Werf, K. O.; de Grooth, B. G.; Greve J.; Hinterdorfer, P.; Gruber, H. J.; Schindler, H.; van Kyook Y.; Figdor C. G. ,Simultaneous height and adhesion imaging of antibody antigen interactions by atomic force microscopy. Biophys. J. 1998, 57, 2220-2228.

47. T. A. Land, A. J. Malkin, Yu.G Kuznetsov, A. McPherson, and J. J. De Yoreo, Mechanisms of Protein Crystal Growth: An Atomic Force Microscopy Study of Canavalin Crystallization: Phys. Rev. Lett. 75, 2774–2777 (1995)

48. Ferry Kienberger , Lilia A. Chtcheglova , Andreas Ebner , Theeraporn Puntheeranurak ,Hermann J. Gruber, Peter Hinterdorfer, 2007,Single-Molecule Studies on Cells and Membranes Using the Atomic Force Microscope, page 101, springer, DOI-10.1007/978-3-540-37319-3_4

49. Sunyoung Lee, Jelena Mandic, and Krystyn J. Van Vliet , Chemomechanical mapping of ligand–receptor binding kinetics on cells, PNAS, June 5, 2007,vol. 104,no. 23,9609–9614

50. Chtcheglova, Lilia,Wildling, Linda, Waschke, Jens,Drenckhahn, Detlev, Hinterdorfer, Peter; AFM Functional Imaging on Vascular Endothelial Cells: Journal of Molecular Recognition 23, 6 (2010) 589,DOI : 10.1002/jmr.105.

51. Ebner A, Wildling L, Zhu R, Rankl C, Haselgrubler T,Hinterdorfer P and Gruber H J 2008; Functionalization of probe tips and supports for single molecule recognition force microscopy Top. Curr. Chem. 285 29–76

52. Peter Hinterdorfer & Yves F Dufrêne ,April 2006,Detection and localization of single molecular recognition events using atomic force microscopy, Nature methods, 1548-7091, volume 3, Issue 5, page 347.

53. Riener CK, Stroh CM, Ebner A, Klampfl C, Gall AA, Romanin C, Lyubchenko YL,Hinterdorfer P, Gruber HJ (2003) Anal Chim Acta 479:59 56.

54. Riener CK, Kienberger F, Hahn CD, Buchinger GM, Egwim IOC, Haselgrubler T,Ebner A, Romanin C, Klampfl C, Lackner B (2003) Anal Chim Acta 497:101

55. Raab A, Han W, Badt D, Smith-Gill SJ, Lindsay SM, Schindler H, Hinterdorfer ,(1999) Nat Biotechnol 17:902

56. Hongda Wang, Ralph Bash, Jiya G. Yodh, Gordon L. Hager, D. Lohr, and Stuart M. Lindsay,Glutaraldehyde Modified Mica: A New Surface for Atomic Force Microscopy of Chromatin: 2002, Biophysical Journal Volume 83 December 2002 3619 –3625

57. Wang, H., L. Obenauer-Kutner, ., S. M. Lindsay. 2008. Imaging glycosylation. J. Am. Chem. Soc. 130:8154–8155.

58. Bash, R., H. Wang, ., D. Lohr. 2006. AFM imaging of protein movements: histone H2A-H2B release during nucleosome remodeling. FEBS Lett. 580:4757–4761.

59. Lin, L., H. Wang, ., S. Lindsay. 2006. Recognition imaging with a DNA aptamer. Biophys. J. 90:4236–4238.

60. Christian K. Riener, Ferry Kienberger,Christoph D. Hahn, Gerhard M. Buchinger, Innocent O.C. Egwima,Thomas Haselgrübler, Andreas Ebner, Christoph Romanin, Christian Klampfl, Bernd Lackner, Heino Prinz, Dieter Blaas, Peter Hinterdorfer, Hermann J. Gruber, Heterobifunctional crosslinkers for tethering single ligand molecules to scanning probes. Analytica Chimica Acta 497 (2003) 101–114.

61. Stroh CM, Ebner A, Geretschlager M, Freudenthaler G, Kienberger F, Kamruzza-han ASM, Smith-Gill SJ, Gruber HJ, Hinterdorfer P (2004) ,Simultaneous Topography and Recognition Imaging Using Force Microscopy :Biophysical Journal Volume 87 September 2004 1981–1990

62. S.scheuring, D.J.Muller,P.Ringler, J.B.Heymann and A.Engel.Imaging streptavidin 2D crystals on biotinylated lipid monolayers at high resolution with the atomic force microscope : Journal of Microscopy, Vol. 193, Pt 1, January 1999, pp. 28–35

63. Dionne C. G. Klein,Cordula M. Stroh,Henriette Jensenius,Maarten van Es,A. S. M. Kamruzzahan,Amalia Stamouli,Hermann J. Gruber,Tjerk H. Oosterkamp,and Peter Hinterdorfer ,Covalent Immobilization of Single proteins on mica for molecular recognition force microscopy,CHEMPHYSCHEM 2003, 4, 1367 ± 1371 DOI: 10.1002/cphc.200300844.

64. Luda S. Shlyakhtenko,Alexander A. Gall, Jeffrey J. Weimer,David D. Hawn, and Yuri L. Lyubchenko, Atomic Force Microscopy Imaging of DNA Covalently Immobilized on a Functionalized Mica Substrate: 1999, Biophysical Journal ,Volume 77, July 1999 568 –576

65. Imaging DNA molecules chemically bound to a mica surface :S.M.Lindsay, Y.L.Lyubchenko,A.GaI, S.Shlyakhtenko and E.ILarington 1992, Proc. SPIE 1639, 84 (1992)

66. Hiroshi Okusa, Kazue Kurihara, Toyoki Kunitake , Chemical Modification Of Molecularly Smooth Mica Surface And Protein Attachment,Langmuir, 1994, 10 (10), pp 3577–3581

67. The Control Of Vascular Integrity By Endothelial Cell Junctions: Molecular Basis And Pathological Implications by :Elisabetta Dejana ,Elisabeth Tournier-Lasserve , Brant M. Weinstein

68. Calcium –dependent cell-cell adhesion molecules(cadherins):subclass specificities and possible involvement of actin bundles : Hirano S, Nose A, Hatta K, Kawakami A, Takeichi M. 1987

68. Vincent PA, Xiao K, Buckley KM, Kowalczyk AP. 2004. VE-cadherin: adhesion at arm's length. Am. J. Physiol. Cell Physiology. : 2004

70. Baumgartner W, Schütz GJ, Wiegand J, Golenhofen N, Drenckhahn D. 2003. Cadherin function probed by laser t weezer and single molecule fluorescence in vascular endothelial cells. J. Cell Sci. 116: 1001-1011.

71. Chemomechanical mapping of ligand –receptor binding kinetics on cells. : Sunyoung Lee, Jelena Mandic and Krystyn J. Van Vliet.

72. Waltenberger J, Claesson-Welsh L, Siegbahnll A, Shibuya M, Heldin C-H(1994)JBiolChem269:26988 –26995

73. Scheidegger P, Weiglhofer W, Suarez S, Console S, Waltenberger J, Pepper MS, Jaussi R, Ballmer-Hofer K (2001) Biochem J 353:569–578.

74. Skou,J.C. Biochem. Biophys. Acta 1957, 1000, 439-446

75. The Corsini Encyclopedia of Psychology and Behavioral Science, Volume 4 : By W. Edward Craighead, Charles B. Nemeroff

76. Localization of Na+ - K+ ATPases in Quasi-Native cell membranes: by Junguang Jiang, Xiao Hao, MIngjun Cai,Yuping Shan, Zhiyong Tang,Hongda Wang

77. Preparation of cell membranes for high resolution imaging by AFM: by Hongda

78. " The structure of DNA": J.D. Watson , F.H.C Crick 1953

79. "Crystal structure of the nucleosome core particle at 2.8 A resolution". Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ (September 1997). Nature 389 (6648): 251–60.

80. kaur , Qiang fu, Alexander Fuhrmann,2011. Antibody Unfolding and Metastable-State Binding in Force Spectroscopy and Recognition Imaging

81. Stroh, C., H. Wang, ., P. Hinterdorfer. 2004. Single-molecule recognition I maging microscopy. Proc. Natl. Acad. Sci. USA. 101:12503.

82. Fuhrmann, A. 2010. Force spectroscopy, from single molecules to whole cells : Refined procedures of data analysis. PhD thesis. Arizona State University.

83. Jagadish, K., and J. A. Camarero. 2010. Cyclotides, a promising molecular scaffold for peptide-based therapeutics. Biopolymers. 94:611–616.

84. Griffith, F, 1928, The Significance of Pneumococcal Types, The Journal of hygiene. Volume 27, Issue 02, page 113.

85. Avery, O. T., 02/1944, Studies On The Chemical Nature Of The Substance Inducing Transformation Of Pneumococcal Types, The Journal of experimental medicine, Volume 79, Issue 02, page 137.

86. Watson and Crick , 04/1993,Genetical Implications Of The Structure Of Deoxyribonucleic Acid,JAMA : the journal of the American Medical Association, Volume 269, Issue 15, page 1967.

87. Franklin, R. E., 09/1953, The Structure Of Sodium Thymonucleate Fibres. I. The Influence Of Water Content, Acta crystallographica, Volume 6, Issue 8, page 673.

88. http://www.accessexcellence.org/RC/VL/GG/structure.php

89. http://www.mun.ca/biology/scarr/iGen3_02-14.html

90. Jones, PL, 1998, Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription, Nature genetics, Volume 19, Issue 2, page 187

91. Antequera, F. ,08/2003, Structure, function and evolution of CpG island promoters, Cellular and molecular life sciences : CMLS, Volume 60, Issue 8, page 1647

92. Razin, A., 09/1998,CpG methylation, chromatin structure and gene silencing---a three-way connection, The EMBO journal, Volume 17, Issue 17, page 4905

93. Wyatt, GR, 1951,Recognition and estimation of 5-methylcytosine in nucleic acids,Biochemical journal, Volume 48, Issue 5, page 581

94. Arber, W, 1962, Host specificity of DNA produced by Escherichia coli. I. Host controlled modification of bacteriophage lambda., Journal of molecular biology. Volume 5 ,page 18.

95. Smith, H , 07/1970 ,A Restriction enzyme from Hemophilus influenzae 1I. Purification and general properties, Journal of molecular biology, Volume 51, Issue 2, page 379.

96. C. Waalwijk, R.A. Flavell, 1978 ,MspI, an isoschizomer of HpaII which cleaves both unmethylated and methylated HpaII sites, Nucleic acids research, Volume 5, Issue 9, page 3231

97. Frommer, M , 1992 ,A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands, PNAS : Proceedings of the National Academy of Sciences, Volume 89, Issue 5, page 1827.

98. Venter, J. C., 02/2001,The Sequence of the Human Genome, Science (New York, N.Y.), Volume 291, Issue 5507, page 1304.

99. Strichman-Almashanu, L. Z., 03/2002, A Genome-Wide Screen for Normally Methylated Human CpG Islands That Can Identify Novel Imprinted Genes, Genome research, Volume 12, Issue 4, page 543

100. Klimasauskas, Saulius, 01/1994,Hhal methyltransferase flips its target base out of the DNA helix, Cell (Cambridge), Volume 76, Issue 2, page 357.

101. Steven S. Smith, Bruce E. Kaplan, Lawrence C. Sowers and Edward M. Newman, May 1992,Mechanism of human methyl-directed DNA

methyltransferase and the fidelity of cytosine methylation, PNAS, Volume 89, page 4744-4748.

102. Sally H Cross and Adrian P Bird, June 1995,CpG islands and genes,Current opinions in genetics and development, Volume 5, Issue 3, Pages 309-314.

103. Bird, Adrian P. , 05/1986 ,CpG-rich islands and the function of DNA methylation, Nature (London), Volume 321, Issue 6067, page 209.

104. Antequera, F, 1993,Number of CpG islands and genes in human and mouse, PNAS : Proceedings of the National Academy of Sciences, Volume 90, Issue 24, page 11995.

105. Antequera, Francisco , 05/2001 ,CpG Islands and DNA Methylation, Encyclopedia of Life Sciences.

106. Gonzalgo, ML ,1997,Mutagenic and epigenetic effects of DNA methylation, Mutation research. Reviews in mutation research, Volume 386, Issue 2, page 107

107. Tazi, J, 03/1990,Alternative chromatin structure at CpG islands, Cell (Cambridge), Volume 60, Issue 6, page 909

108. Antequera, Francisco , 08/1990 ,High levels of De Novo methylation and altered chromatin structure at CpG islands in cell lines, Cell (Cambridge),Volume 62, Issue 3, page 503

109. Kratky and G. Porod, Red. Trau. Chim.,68,1106 (1949)

110. Frontali, C., Dore, E., Ferrauto, A., Gratton, E., Bettini, A., Pozzan, M.R. and Valdevit, E. (1979), An absolute method for the determination of the persistence length of native DNA from electron micrographs. Biopolymers, 18(6), 1353-1373.

111. Hansma, H.G., Kim, K.J., Laney, D.E., Garcia, R.A., Argaman, M., Allen, M.J. and Parsons, S.M. (1997) Properties of biomolecules measured from atomic force microscope images: a review. J. Struct. Biol., 119(2), 99-108.

112. Rivetti, C., Guthold, M. and Bustamante, C. (1996) Scanning force microscopy of DNA deposited onto mica: equilibration versus kinetic trapping studied by statistical polymer chain analysis. J. Molec. Biol., 264(5), 919-932.

113. Marek, J., Demjenova, E., Interactive measurement and characterization of DNA molecules by analysis of AFM images. Cytometry A, 63(2), 87-93.

114. Rivetti, C., Guthold, M. and Bustamante, C. (1996) Scanning force microscopy of DNA deposited onto mica: equilibration versus kinetic trapping studied by statistical polymer chain analysis. J. Molec. Biol., 264(5), 919-932.

115. Smith, S.B., Finzi, L. and Bustamante, C. (1992) Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads. Science, 258(5085), 1122-1126.

116. Dietmar Porschke. (1990) Persistence length and bending dynamics of DNA from electrooptical measurements at high salt concentrations

117. Michael Rubinstein , Ralph H.Colby (2003) Polymer Physics.

118. Teraoka, (2002)Polymer solutions: An introduction to physical properties.

119. Hagerman, P. J. Hodges- Garcia, Y. a. (1995). Investigation of the influence of cytosine methylation on DNA flexibility, The Journal of biological chemistry, Volume 270, Issue 1, page 197.

120. Nathan, D. a. (2002). Bending and Flexibility of Methylated and Unmethylated EcorI DNA, Journal of molecular biology, Volume 316, Issue 1, page 7.

121. Paul A. Wiggins et al, 2006,High flexibility of DNA on short length scales probed by atomic force microscopy. Nature Nanotechnology, Volume 1.

122. Static and dynamic light Scattering on DNA solution, http://oldweb.ct.infn.it/~zimbone/LSO/softmatter/Dynamic%20Scattering.pdf

123. Shashkov, SN, The study of DMSO/water and DPPC/DMSO/water system by means of the X-ray, neutron small-angle scattering, calorimetry and IR spectroscopy, 1999,Physica. B, Condensed matter, Volume 271, Issue 1-4, page 184.

124. Rubinstein, P, 1995, Processing and cryopreservation of placental/umbilical cord blood for unrelated bone marrow reconstitution, PNAS : Proceedings of the National Academy of Sciences, Volume 92, Issue 22, page 10119 .

125. Gu, Qun, 01/2006, DNA nanowire fabrication, Nanotechnology,Volume17, Issue 1, page R14 Stoltenberg, Randall M., 06/2004,DNA-Templated Nanowire Fabrication ,Biomedical microdevices, Volume 6, Issue 2, page 105.

126. Paul A Wade, 2005, Switching off methylated DNA , Nature genetics, Volume 37, Issue 3, Page 212.

127. Razin, A.,CpG methylation, chromatin structure and gene silencing---a three-way connection, The EMBO journal, Volume 17,Issue 17, Page 4905.

128. Choy et al, 2010, DNA methylation increases Nucleosome Compaction and Rigidity, Journal of the American Chemical Society, Volume 132,Issue 6, Page 1782.

129. Hansen, JC,1991 ,The mechanism of nucleosome assembly onto oligomers of the sea urchin 5 S DNA positioning sequence, The Journal of biological chemistry, Volume 266,Issue 7, Page 4276.

130. Widom, J, 2001,Role of DNA sequence in nucleosome stability and dynamics, Quarterly reviews of biophysics, Volume 34, Issue 3, page 269

131. Lowary, Widom,New DNA Sequence Rules for High Affinity Binding to Histone Octamer and Sequence-directed Nucleosome Positioning, J. Mol. Biol. (1998) 276, Pages 19-42

132. KARYMOV, M. A.,2001, DNA methylation-dependent chromatin fiber compaction invivo and in vitro: requirement for linker histone, The FASEB journal, Volume 15, Issue 14, page 2631

APPENDIX A

MATLAB CODE FOR ANALYZING AFM IMAGES FOR PERSISTENCE LENGTH AND CONTOUR LENGTH MEASUREMENTS

dydx:

```
function dy=dydx(y,x);
% dy=dydx(y,x);
% This function calculates the derivative of a function y given at the
% points x
narg=nargin;
df=diff(y);
if narg>1
   df=df./diff(x);
end
dy=mean([[df;df(end)] [df(1);df]],2);
```

interpsel:

```
function [tx,ty]=interpsel(x, y, step)
x=x(:)';
y=y(:)';
t=cumsum([0 sqrt(diff(x).^2+diff(y).^2)]);
A=spline(t,[x;y],t(1):step:t(end)+step/2);
tx=A(1,:);
ty=A(2,:);
```

niceplot:

```
function h=niceplot(fun,lab,leg,varargin)
if nargin < 3
   error('Not enough input arguments');
end
h.fig = figure;
set(h.fig, 'DefaultTextFontSize', 16);
set(h.fig, 'DefaultAxesFontSize', 16);
%set(h.fig, 'DefaultAxesFontName', 'Times New Roman');
```

```
%set(h.fig, 'DefaultTextFontName', 'Times New Roman');
set(h.fig, 'Units', 'centimeters');
pos = get(h.fig, 'Position');
pos(3) = 8;
pos(4) = 6;
set(h.fig, 'Position', pos);
set(h.fig, 'PaperType', 'a4letter');
set(h.fig, 'PaperPositionMode', 'Auto');
set(h.fig, 'Color', 'None');
set(h.fig, 'InvertHardcopy', 'Off');
h.plot=feval(fun,varargin{:});

set(h.plot,'LineWidth',3);
set(gca,'FontSize',16,'FontWeight','bold','LineWidth',2, 'Color', 'None');

if ~iscell(lab)
   error('argument label mut be a cell array');
end

if length(lab) < 3
   lab{end+1:3}='';
end
h.xlabel=xlabel(lab{1});
h.ylabel=ylabel(lab{2});
set([h.xlabel h.ylabel],'FontSize',20);
h.title=title(lab{3});
set(h.title,'Fontsize',24);
```

```
if ~isempty(leg)
   h.legend=legend(leg);
   set(h.legend,'Fontsize',14);
else
   h.legend=nan;
end
```

pfit:

```
function p=pfit(x,y,n)
%
% Find polynom of degree n which fits in a least mean squere sence
% through the points defined by x and y
%
% p=pfit(x,y,n)
%
% Input x: Vector with the x-coordinate of the points
%       y: Vector with the y-coordinate of the points
%       n: Scalar with the degree of the fit polynom
%          or a vector with the degrees which should
%          be considerd in the fit
%
% Examples
%
% Finds a 3rd order polynom through the points (x,y):
% x=rand(1,10);y=rand(1,10);
% p=pfit(x,y,3)
%
%
% fit a polynom of 3rd order but no x term
```

```
% x=rand(1,10);y=rand(1,10);
% p=pfit(x,y,[0 2 3]);
%
if ~isvector(x)
  error('x must be a vector');
end
x=x(:);
y=y(:);
if size(x,1)~=size(y,1)
  error('x and y must have the same number of rows');
end
% Koeffizientenmatrix erstellen
if length(n)==1
  anz=n;
else
  anz=max(n);
end
A=[ones(length(x),1),repmat(x,[1,anz])];
A=cumprod(A,2);
if length(n)==1
  p=A\y;
else
  tmp=A(:,n+1)\y;
  p=zeros(anz+1,1);
  p(n+1)=tmp;
end
p=flipud(p).';
```

plots:

```
function plots
pfad=uigetdir;
if isequal(pfad,0)
   return
end
s=dir(fullfile(pfad, '*.mat'));
atx=cell(0,1);
aty=atx;
rtx=atx;
rty=atx;
otx=atx;
oty=atx;
for k=1:length(s)
   load(fullfile(pfad, s(k).name),'ok','ring','tx','ty','step','nmperpx')
   if ~exist('tx','var')
      continue;
   end
   ind=1:length(tx);
   ok=ok(ind);
   ring=ring(ind);
   tx=cellfun(@(x)(x*nmperpx),tx,'UniformOutput',false);
   ty=cellfun(@(x)(x*nmperpx),ty,'UniformOutput',false);
   atx=cat(1,atx,tx(ok));
   aty=cat(1,aty,ty(ok));
   rtx=cat(1,rtx,tx(ok & ring));
   rty=cat(1,rty,ty(ok & ring));
```

```
    otx=cat(1, otx, tx(ok & ~ring));
    oty=cat(1, oty, ty(ok & ~ring));
end
% Verteilung der DNA Lï¿½ngen
tmp=cellfun(@length,atx,'UniFormOutput',true)*2.5;
%x=linspace(0,250,250);
%[aly, alx]=ksdensity(tmp, x);
[aly, alx]= hist(tmp);
%aly=aly*length(tmp);
tmp=cellfun(@length,otx,'UniFormOutput',true)*2.5;
%[oly, olx]=ksdensity(tmp, x);
[oly, olx]=hist(tmp,alx);
%oly=oly*length(tmp);
tmp=cellfun(@length,rtx,'UniFormOutput',true)*2.5;
%[rly, rlx]=ksdensity(tmp, x);
[rly,rlx]=hist(tmp, alx);
%rly=rly*length(tmp);
figure;
%bar(alx,aly,rlx,rly,olx,oly);
bar(alx, [aly(:) rly(:) oly(:)]);
title('DNA Length Distribution');
xlabel('DNA Length');
ylabel('Occurence');
latx=length(atx);
lotx=length(otx);
lrtx=length(rtx);
%seps=[2 4 12];
seps=2:2:40;
```

```
lsp=length(seps);
dphi=cell(lsp,1);
dr = dphi;
G=dphi;
drphi=dphi;
drr=dphi;
Gr=G;
dophi=dphi;
dor=dphi;
Go=G;
for k=1:latx
   dx = atx{k}(3:end)-atx{k}(1:end-2);
   dy = aty{k}(3:end)-aty{k}(1:end-2);
   phi = unwrap(atan2(dy,dx));
   for l=1:lsp
      dphi{l}=cat(2,dphi{l},phi(1:end-seps(l))-phi(1+seps(l):end));
      tmpx=atx{k}(1+seps(l):end)-atx{k}(1:end-seps(l));
      tmpy=aty{k}(1+seps(l):end)-aty{k}(1:end-seps(l));
      dr{l} = cat (2,dr{l}, tmpx.^2+tmpy.^2);
   end
   if k<=lrtx
      dx = rtx{k}(3:end)-rtx{k}(1:end-2);
      dy = rty{k}(3:end)-rty{k}(1:end-2);
      phi = unwrap(atan2(dy,dx));
      for l=1:lsp
         drphi{l}=cat(2,drphi{l},phi(1:end-seps(l))-phi(1+seps(l):end));
         tmpx=rtx{k}(1+seps(l):end)-rtx{k}(1:end-seps(l));
         tmpy=rty{k}(1+seps(l):end)-rty{k}(1:end-seps(l));
```

```
        drr{l} = cat (2,drr{l}, tmpx.^2+tmpy.^2);
      end
    end
    if k<=lotx
      dx = otx{k}(3:end) - otx{k}(1:end-2);
      dy = oty{k}(3:end) - oty{k}(1:end-2);
      phi = unwrap(atan2(dy,dx));
      for l=1:lsp
        dophi{l}=cat(2,dophi{l},phi(1:end-seps(l))-phi(1+seps(l):end));
        tmpx=otx{k}(1+seps(l):end)-otx{k}(1:end-seps(l));
        tmpy=oty{k}(1+seps(l):end)-oty{k}(1:end-seps(l));
        dor{l} = cat (2,dor{l}, tmpx.^2+tmpy.^2);
      end
    end
end
ind = seps'*2.5;
meanrsq  = cellfun(@(x)mean(x), dr, 'UniformOutput', true);
stdrsq = cellfun(@(x)std(x)/sqrt(length(x)), dr, 'UniformOutput', true);
meanrrsq = cellfun(@(x)mean(x), drr, 'UniformOutput', true);
stdrrsq = cellfun(@(x)std(x)/sqrt(length(x)), drr, 'UniformOutput', true);
meanorsq = cellfun(@(x)mean(x), dor, 'UniformOutput', true);
stdorsq = cellfun(@(x)std(x)/sqrt(length(x)), dor, 'UniformOutput', true);
meancost  = cellfun (@(x)mean(cos(abs(x))), dphi, 'UniformOutput', true);
stdcost  = cellfun (@(x)std(cos(abs(x)))/sqrt(length(x)), dphi, 'UniformOutput', true);
wa=1./stdcost;
wa = wa / mean(wa);
meancostr = cellfun(@(x)mean(cos(abs(x))), drphi, 'UniformOutput', true);
stdcostr = cellfun(@(x)std(cos(abs(x)))/sqrt(length(x)), drphi, 'UniformOutput', true);
```

```
wr=1./stdcostr;
wr=wr/mean(wr);
meancosto = cellfun(@(x)mean(cos(abs(x))), dophi, 'UniformOutput', true);
stdcosto = cellfun(@(x)std(cos(abs(x)))/sqrt(length(x)), dophi, 'UniformOutput', true);
wo=1./stdcosto;
wo=wo/mean(wo);
f=inline('exp(-x/(2*xi))','xi','x');
xi=nlinfit(ind, meancost.*sqrt(wa),@(xi,x)sqrt(wa).*f(xi,x), 20);
xir=nlinfit(ind, meancostr.*sqrt(wr),@(xi,x)sqrt(wr).*f(xi,x), 20);
xio=nlinfit(ind, meancosto.*sqrt(wo),@(xi,x)sqrt(wo).*f(xi,x), 20);
g=inline('4*xi*(x+2*xi*(exp(-x/(2*xi))-1))','xi','x');
figure;
plot(ind, meanrsq,'.', ind, meanrrsq,'.',...
   ind, meanorsq,'.');
hold on;
plot(ind, g(xi, ind),'--',ind, g(xir, ind), '--', ind, g(xio, ind),'--');
xlabel('Separation, L(nm)');ylabel('<(R_{s,s+L})^2>');
legend(num2str([xi;xir;xio]));
figure;
plot(ind, meancost,'.', ind, meancostr,'.',...
   ind, meancosto,'.');
hold on;
plot(ind, f(xi, ind),'--', ind, f(xir,ind),'--', ind, f(xio, ind),'--');
xlabel('Separation, L(nm)'); ylabel('<cos \theta_{s,s+L}>');
legend(num2str([xi;xir;xio]));
theta=linspace(0.1,1.5,100);
sseps=[2 4 12];
figure;
```

```
clf;
hold on;
c='bgrcm';
ls=['+o*.xsd^v><ph'];
for l=1:length(sseps);
   G{l} = ksdensity(abs(dphi{sseps(l)}),theta,'support','positive');
   Go{l} = ksdensity(abs(dophi{sseps(l)}), theta,'support','positive');
   if ~isempty(drphi{sseps(l)})
      Gr{l} = ksdensity(abs(drphi{sseps(l)}),theta,'support','positive');
   else
      Gr{l} = nan;
   end
   p=pfit(theta, -log(G{l}),[0 2]);
   if ~isnan(Gr{l})
      pr=pfit(theta, -log(Gr{l}),[0 2]);
   else
      pr=[nan nan];
   end
   po=pfit(theta, -log(Go{l}),[0 2]);
   plot(theta,-log(G{l}),[c(l),ls(1)],theta,-log(Go{l}),[c(l),ls(2)],...
      theta,-log(Gr{l}),[c(l),ls(3)]);
end
xlabel('\theta','Interpreter','Tex');
ylabel('-log(G(\theta;L))');
return
seps=[3 6 20];
lsp=length(seps);
dr=cell(lsp,1);
```

```
K=dr;
drr=dr;
Kr=K;
dor=dr;
Ko=K;
xk=K;
for k=1:latx
   for l=1:lsp
      dx = atx{k}(1:end-seps(l))-atx{k}(1+seps(l):end);
      dy = aty{k}(1:end-seps(l))-aty{k}(1+seps(l):end);
      dr{l}=cat(2,dr{l},sqrt(dx.^2+dy.^2));
   end
   if k<=lrtx
      for l=1:lsp
         dx = rtx{k}(1:end-seps(l))-rtx{k}(1+seps(l):end);
         dy = rty{k}(1:end-seps(l))-rty{k}(1+seps(l):end);
         drr{l}=cat(2,drr{l},sqrt(dx.^2+dy.^2));
      end
   end
   if k<=lotx
      for l=1:lsp
         dx = otx{k}(1:end-seps(l))-otx{k}(1+seps(l):end);
         dy = oty{k}(1:end-seps(l))-oty{k}(1+seps(l):end);
         dor{l}=cat(2,dor{l},sqrt(dx.^2+dy.^2));
      end
   end
end
```

```
figure;
clf;
hold on;
c='rgbcm';
for l=1:lsp
    [K{l}, xk{l}] = ksdensity(dr{l});
    [Ko{l}] = ksdensity(dor{l},xk{l});
    if ~isempty(drr{l})
        [Kr{l}] = ksdensity(drr{l},xk{l});
    else
        Kr{l} = nan;
    end
    plot(xk{l}/seps(l), K{l}, c(l), xk{l}/seps(l), Ko{l}, [c(l),'--'],...
        xk{l}/seps(l), Kr{l},[c(l),'-.']);
end
xlabel('R/L','Interpreter','Tex');
ylabel('K(R;L)');
```

plots:

```
function plots2
pfad=uigetdir;
if isequal(pfad,0)
    return
end
s=dir(fullfile(pfad,'*.mat'));
atx=cell(0,1);
aty=atx;
rtx=atx;
rty=atx;
```

```
otx=atx;
oty=atx;
for k=1:length(s)
    load(fullfile(pfad,s(k).name),'ok','ring','tx','ty','step','nmperpx')
    if ~exist('tx','var')
        fullfile(pfad,s(k).name)
        continue;
    end
    ind=1:length(tx);
    ok=ok(ind);
    ring=ring(ind);
    tx=cellfun(@(x)(x*nmperpx),tx,'UniformOutput',false);
    ty=cellfun(@(x)(x*nmperpx),ty,'UniformOutput',false);
    atx=cat(1,atx,tx(ok));
    aty=cat(1,aty,ty(ok));
    rtx=cat(1,rtx,tx(ok & ring));
    rty=cat(1,rty,ty(ok & ring));
    otx=cat(1, otx, tx(ok & ~ring));
    oty=cat(1, oty, ty(ok & ~ring));
end
seps=(2:2:40)';
sepsg=[2 4 12];
sepsk=[3 6 20];
c='bgrcmyk';
f=inline('exp(-x/(2*xi))','xi','x');
g=inline('4*xi*(x+2*xi*(exp(-x/(2*xi))-1))','xi','x');
```

```
%out1=dnalyz(atx, aty, seps, sepsg, sepsk);
out2=dnalyz(otx, oty, seps, sepsg, sepsk);
if ~isempty(rtx)
   out3=dnalyz(rtx, rty, seps, sepsg, sepsk);
end
figure;
for k=1:3
   if exist(sprintf('out%i',k),'var')
      eval(sprintf('pplot({"","",""},{}, out%i.lx, out%i.ly,"%s");', k, k,c(k)));
      %k
      eval(sprintf('out%i.len',k));
      eval(sprintf('out%i.dnas',k));
      disp('');
   end
   hold on;
end
pplot({'DNA length [nm]','occurence', ''},{},nan,nan);
saveas(gcf,fullfile(pfad,'dnalen'),'fig');
figure;
for k=1:3
   if exist(sprintf('out%i',k),'var')
      eval(sprintf(['niceplot(@errorbar,{"","",""},{},out%i.L,',...
         'out%i.mct, out%i.sct,"%s.");'], k, k, k,c(k)));
   end
   hold on;
end
```

```
for k=1:3
   if exist(sprintf('out%i',k),'var')
      eval(sprintf(['pplot({"","",""},{},out%i.L,',...
         'f(out%i.xi, out%i.L),"%s--");'], k, k,k,c(k)));
      k
      eval(['out',num2str(k),'.xi']);
   end
end
pplot({'Separation, L [nm]','<cos\theta_{s,s+L}>', ''},{},nan,nan);
saveas(gcf,fullfile(pfad,'cost'),'fig');
figure;
for k=1:3
   if exist(sprintf('out%i',k),'var')
      eval(sprintf(['niceplot(@errorbar,{"","",""},{},out%i.L,',...
         'out%i.mr2, out%i.sr2,"%s.");'], k, k,k,c(k)));
   end
   hold on;
end
for k=1:3
   if exist(sprintf('out%i',k),'var')
      eval(sprintf(['pplot({"","",""},{},out%i.L,',...
         'g(out%i.xi, out%i.L),"%s--");'], k, k,k,c(k)));
   end
end
pplot({'Separation, L [nm]', '<R_{s,s+L}^2> [nm^2]',''},...
   {}, nan, nan);
saveas(gcf,fullfile(pfad,'rsq'),'fig');
function out=dnalyz(tx,ty,seps, sepsg, sepsk)
```

```
ltx=length(tx);
% Distribution of length
tmp=cellfun(@length,tx,'UniFormOutput',true)*2.5;
tmp=tmp(tmp>0);
out.dnas=length(tmp);
out.len=sum(tmp);
%[out.ly,out.lx]=smoothhist(tmp,30);
[out.ly, out.lx]=ksdensity(tmp);
lsp=length(seps);
out.dphi=cell(lsp,1);
out.dr = out.dphi;
for k=1:ltx
   dx = tx{k}(3:end)-tx{k}(1:end-2);
   dy = ty{k}(3:end)-ty{k}(1:end-2);
   phi = unwrap(atan2(dy,dx));
   for l=1:lsp
      tmpphi=phi(1:end-seps(l))-phi(1+seps(l):end);
      out.dphi{l}=cat(2,out.dphi{l},tmpphi);
      tmpx=tx{k}(1+seps(l):end)-tx{k}(1:end-seps(l));
      tmpy=ty{k}(1+seps(l):end)-ty{k}(1:end-seps(l));
      out.dr{l} = cat (2,out.dr{l}, tmpx.^2+tmpy.^2);
   end
end
m=@(x)mean(x);
s=@(x)std(x)/sqrt(length(x));
out.mr2 = cellfun(m, out.dr,'UniformOutput', true);
out.sr2 = cellfun(s, out.dr, 'UniformOutput', true);
```

```
c=@(x)(cos(x));
out.mct = cellfun(@(x)m(c(x)), out.dphi, 'UniformOutput', true);
out.sct = cellfun(@(x)s(c(x)), out.dphi, 'UniformOutput', true);
w=1./out.sct;
w=w/mean(w);
out.L = seps*2.5;
f=inline('exp(-x/(2*xi))','xi','x');
ind=isfinite(out.mct);
if all(ind)
   out.xi=nlinfit(out.L, out.mct.*sqrt(w),...
      @(xi,x)sqrt(w).*f(xi,x), 58);
else
   out.xi=nlinfit(out.L(ind), out.mct(ind), f, 58);
end
out.G=cell(length(sepsg),1);
out.t=out.G;
out.p=out.G;
for l=1:length(sepsg);
   [out.G{l}, out.t{l}] = smoothhist(abs(out.dphi{sepsg(l)}));
   out.p{l}=pfit(out.t{l}, -log(out.G{l}),[0 2]);
end
```

plots_bp :

```
function plots_bp
pfad=uigetdir;
if isequal(pfad,0)
   return
end
```

```
s=dir(fullfile(pfad, '*.mat'));
atx=cell(0,1);
aty=atx;
rtx=atx;
rty=atx;
otx=atx;
oty=atx;
for k=1:length(s)
    load(fullfile(pfad, s(k).name),'ok','ring','tx','ty','step','nmperpx')
    if ~exist('tx','var')
        continue;
    end
    ind=1:length(tx);
    ok=ok(ind);
    ring=ring(ind);
    tx=cellfun(@(x)(x*nmperpx),tx,'UniformOutput',false);
    ty=cellfun(@(x)(x*nmperpx),ty,'UniformOutput',false);
    [tx]
    atx=cat(1,atx,tx(ok));
    aty=cat(1,aty,ty(ok));
    rtx=cat(1,rtx,tx(ok & ring));
    rty=cat(1,rty,ty(ok & ring));
    otx=cat(1, otx, tx(ok & ~ring));
    oty=cat(1, oty, ty(ok & ~ring));
end
% Verteilung der DNA Lï¿½ngen
tmp=cellfun(@length,atx,'UniFormOutput',true)*2.5;
```

```
%x=linspace(0,250,250);

%[aly, alx]=ksdensity(tmp, x);

[aly, alx]= hist(tmp);

%aly=aly*length(tmp);

tmp=cellfun(@length,otx,'UniFormOutput',true)*2.5;

%[oly, olx]=ksdensity(tmp, x);

[oly, olx]=hist(tmp,alx);

%oly=oly*length(tmp);

tmp=cellfun(@length,rtx,'UniFormOutput',true)*2.5;

%[rly, rlx]=ksdensity(tmp, x);

[rly,rlx]=hist(tmp, alx);

%rly=rly*length(tmp);

figure;

%bar(alx,aly,rlx,rly,olx,oly);

bar(alx, [aly(:) rly(:) oly(:)]);

title('DNA Length Distribution');

xlabel('DNA Length');

ylabel('Occurence');

[alx]

latx=length(atx);

lotx=length(otx);

lrtx=length(rtx);

[alx]

%seps=[2 4 12];

seps=2:2:40;

lsp=length(seps);

dphi=cell(lsp,1);

dr = dphi;
```

```
G=dphi;
drphi=dphi;
drr=dphi;
Gr=G;
dophi=dphi;
dor=dphi;
Go=G;
for k=1:latx
   dx = atx{k}(3:end)-atx{k}(1:end-2);
   dy = aty{k}(3:end)-aty{k}(1:end-2);
   phi = unwrap(atan2(dy,dx));
   for l=1:lsp
      dphi{l}=cat(2,dphi{l},phi(1:end-seps(l))-phi(1+seps(l):end));
      tmpx=atx{k}(1+seps(l):end)-atx{k}(1:end-seps(l));
      tmpy=aty{k}(1+seps(l):end)-aty{k}(1:end-seps(l));
      dr{l} = cat (2,dr{l}, tmpx.^2+tmpy.^2);
   end
   if k<=lrtx
      dx = rtx{k}(3:end)-rtx{k}(1:end-2);
      dy = rty{k}(3:end)-rty{k}(1:end-2);
      phi = unwrap(atan2(dy,dx));
      for l=1:lsp
         drphi{l}=cat(2,drphi{l},phi(1:end-seps(l))-phi(1+seps(l):end));
         tmpx=rtx{k}(1+seps(l):end)-rtx{k}(1:end-seps(l));
         tmpy=rty{k}(1+seps(l):end)-rty{k}(1:end-seps(l));
         drr{l} = cat (2,drr{l}, tmpx.^2+tmpy.^2);
      end
   end
```

```
    if k<=lotx
        dx = otx{k}(3:end) - otx{k}(1:end-2);
        dy = oty{k}(3:end) - oty{k}(1:end-2);
        phi = unwrap(atan2(dy,dx));
        for l=1:lsp
            dophi{l}=cat(2,dophi{l},phi(1:end-seps(l))-phi(1+seps(l):end));
            tmpx=otx{k}(1+seps(l):end)-otx{k}(1:end-seps(l));
            tmpy=oty{k}(1+seps(l):end)-oty{k}(1:end-seps(l));
            dor{l} = cat (2,dor{l}, tmpx.^2+tmpy.^2);
        end
    end
end
ind = seps'*2.5;
meanrsq  = cellfun(@(x)mean(x), dr, 'UniformOutput', true);
stdrsq = cellfun(@(x)std(x)/sqrt(length(x)), dr, 'UniformOutput', true);
meanrrsq = cellfun(@(x)mean(x), drr, 'UniformOutput', true);
stdrrsq = cellfun(@(x)std(x)/sqrt(length(x)), drr, 'UniformOutput', true);
meanorsq = cellfun(@(x)mean(x), dor, 'UniformOutput', true);
stdorsq = cellfun(@(x)std(x)/sqrt(length(x)), dor, 'UniformOutput', true);
meancost  = cellfun (@(x)mean(cos(abs(x))), dphi, 'UniformOutput', true);
stdcost  = cellfun (@(x)std(cos(abs(x)))/sqrt(length(x)), dphi, 'UniformOutput', true);
wa=1./stdcost;
wa = wa / mean(wa);
meancostr = cellfun(@(x)mean(cos(abs(x))), drphi, 'UniformOutput', true);
stdcostr = cellfun(@(x)std(cos(abs(x)))/sqrt(length(x)), drphi, 'UniformOutput', true);
wr=1./stdcostr;
wr=wr/mean(wr);
```

```
meancosto = cellfun(@(x)mean(cos(abs(x))), dophi, 'UniformOutput', true);

stdcosto = cellfun(@(x)std(cos(abs(x)))/sqrt(length(x)), dophi, 'UniformOutput', true);

wo=1./stdcosto;

wo=wo/mean(wo);

 f=inline('exp(-x/(2*xi))','xi','x');

xi=nlinfit(ind, meancost.*sqrt(wa),@(xi,x)sqrt(wa).*f(xi,x), 20);

xir=nlinfit(ind, meancostr.*sqrt(wr),@(xi,x)sqrt(wr).*f(xi,x), 20);

xio=nlinfit(ind, meancosto.*sqrt(wo),@(xi,x)sqrt(wo).*f(xi,x), 20);

g=inline('4*xi*(x+2*xi*(exp(-x/(2*xi))-1))','xi','x');

figure;

plot(ind, meanrsq,'.', ind, meanrrsq,'.',...

    ind, meanorsq,'.');

hold on;

plot(ind, g(xi, ind),'--',ind, g(xir, ind), '--', ind, g(xio, ind),'--');

xlabel('Separation, L(nm)');ylabel('<(R_{s,s+L})^2>');

legend(num2str([xi;xir;xio]));

figure;

plot(ind, meancost,'.', ind, meancostr,'.',...

    ind, meancosto,'.');

hold on;

plot(ind, f(xi, ind),'--', ind, f(xir,ind),'--', ind, f(xio, ind),'--');

xlabel('Separation, L(nm)'); ylabel('<cos \theta_{s,s+L}>');

legend(num2str([xi;xir;xio]));

theta=linspace(0.1,1.5,100);

sseps=[2 4 12];

figure;

clf;

hold on;
```

```
c='bgrcm';
ls=['+o*.xsd^v><ph'];
for l=1:length(sseps);
   G{l} = ksdensity(abs(dphi{sseps(l)}),theta,'support','positive');
   Go{l} = ksdensity(abs(dophi{sseps(l)}), theta,'support','positive');
   if ~isempty(drphi{sseps(l)})
      Gr{l} = ksdensity(abs(drphi{sseps(l)}),theta,'support','positive');
   else
      Gr{l} = nan;
   end
   p=pfit(theta, -log(G{l}),[0 2]);
   if ~isnan(Gr{l})
      pr=pfit(theta, -log(Gr{l}),[0 2]);
   else
      pr=[nan nan];
   end
   po=pfit(theta, -log(Go{l}),[0 2]);
   plot(theta,-log(G{l}),[c(l),ls(1)],theta,-log(Go{l}),[c(l),ls(2)],...
      theta,-log(Gr{l}),[c(l),ls(3)]);
end
xlabel('\theta','Interpreter','Tex');
ylabel('-log(G(\theta;L))');
return
seps=[3 6 20];
lsp=length(seps);
dr=cell(lsp,1);
K=dr;
drr=dr;
```

```
Kr=K;
dor=dr;
Ko=K;
xk=K;
for k=1:latx
   for l=1:lsp
      dx = atx{k}(1:end-seps(l))-atx{k}(1+seps(l):end);
      dy = aty{k}(1:end-seps(l))-aty{k}(1+seps(l):end);
      dr{l}=cat(2,dr{l},sqrt(dx.^2+dy.^2));
   end
   if k<=lrtx
      for l=1:lsp
         dx = rtx{k}(1:end-seps(l))-rtx{k}(1+seps(l):end);
         dy = rty{k}(1:end-seps(l))-rty{k}(1+seps(l):end);
         drr{l}=cat(2,drr{l},sqrt(dx.^2+dy.^2));
      end
   end
   if k<=lotx
      for l=1:lsp
         dx = otx{k}(1:end-seps(l))-otx{k}(1+seps(l):end);
         dy = oty{k}(1:end-seps(l))-oty{k}(1+seps(l):end);
         dor{l}=cat(2,dor{l},sqrt(dx.^2+dy.^2));
      end
   end
end
figure;
clf;
hold on;
```

```
c='rgbcm';
for l=1:lsp
    [K{l}, xk{l}] = ksdensity(dr{l});
    [Ko{l}] = ksdensity(dor{l},xk{l});
    if ~isempty(drr{l})
        [Kr{l}] = ksdensity(drr{l},xk{l});
    else
        Kr{l} = nan;
    end
    plot(xk{l}/seps(l), K{l}, c(l), xk{l}/seps(l), Ko{l}, [c(l),'--'],...
        xk{l}/seps(l), Kr{l},[c(l),'-.']);
end
xlabel('R/L','Interpreter','Tex');
ylabel('K(R;L)');
```

refinepos:

```
function [rx,ry]=refinepos(bild, x, y,X,Y,step)
t = linspace(-3, 3, 30);
lx = length(x);
rx=nan(lx,1);
ry=rx;
dx = diff(x);
dx = mean([[dx(1);dx],[dx;dx(end)]],2);
dy = diff(y);
dy = mean([[dy(1);dy],[dy;dy(end)]],2);
r = sqrt(dx.^2 + dy.^2);
dx=dx./r;
dy=dy./r;
```

```
for k=1:length(x)
    count = 0;
    while (count < 1)
        normx = x(k) - t * dy(k) * step;
        normy = y(k) + t * dx(k) * step;
        % Caclulate the image along the above line
        A = interp2 (X, Y, bild, normx, normy, '*linear');
        % Find the maximum, use a robust estimation:
        % find the biggest part which is over min(A) + range(A)/3
        % find thru these data a 2nd order polynomial, use the maximum
        % of this function as maximum estimate for the DNA
        mxa=max(A);
        mna=min(A);
        aind = A > mna+(mxa-mna)/3;
        fi = find([0,diff(aind)]);
        [mxfi,fii]=max(diff(fi));
        aind=fi(fii):fi(fii+1)-1;
        if sum(aind)>3
            p=polyfit(t(aind), A(aind),2);
            tmax=-p(2)/(2*p(1));
            rx(k) = x(k) - tmax * dy(k)* step;
            ry(k) = y(k) + tmax * dx(k) * step;
        else
            rx(k)=x(k);
            ry(k)=y(k);
        end
        count = count + 1;
```

```
    end
end
```

Smoothhist:

```
  function [sy,sx]=smoothhist(varargin)
[heights, centers] = hist(varargin{:});
n = length(centers);
w = centers(2) - centers(1);
t = linspace(centers(1)-w/2, centers(end)+w/2, n+1);
p = fix(n/2);
dt = diff(t);
Fvals = cumsum([0,heights.*dt]);
F = spline(t, [0, Fvals, 0]);
sx = linspace(min(varargin{1}), max(varargin{1}),1001)';
sy = ppval(F, sx);
sy = dydx(sy,sx);
```

trackdna:

```
function trackdna
figure;
files=dir('*.mat');
stepnm = 2.5; % Generate a point at every 2.5 nm
datafiles=dir('../Data/080708/*.stp');
nr=cellfun(@(x)str2double(...
    x(find(x=='_',1,'last')+1:find(x=='.',1,'last')-1)),{datafiles.name});
for l=1:length(files)
    load(files(l).name);
    sb=size(bild);
```

```
bild=bild-median(bild(:));
% Do we know the nm per pixel ratio
if ~exist('nmperpx','var')
   % Look if the raw data file can be found
   [p,f,e]=fparts(files(l).name); %#ok<NASGU>
   f=str2double(f);
   f=fullfile('../Data/080708',datafiles(find(nr==f,1,'first')).name);
   s=dir([f,'*']);
   found=cellfun(@(x)strcmpi(x(end-3:end),'.stp'), {s.name});
   if any(found)
      [tmp,bheader]=stpread(f);
   else
      answ=inputdlg(sprintf('Scan Size [nm] of the image in the file %s',...
         files(l).name));
      if isempty(answ)
         return
      end
      answ=answ{1};
      bheader{1}.length_x=str2double(answ)*10;
      if ~isfinite(bheader{1}.length_x)
         [tmp,bheader]=stpread(answ);
         if ~exist('tmp','var') || isempty(tmp)
            return
         end
      end
   end
   nmperpx=bheader{1}.length_x/(10*sb(2));
end
```

```
    step=stepnm/nmperpx;
    ind = cellfun(@(x)~isempty(x),x); %#ok<NODEF>
    x=x(ind);
    y=y(ind); %#ok<NODEF>
    subplot(1,2,1);
    imagesc(bild);
    title(sprintf('File: %s',files(l).name));
    hold on;
    axis equal
    axis tight
    border=plot(nan,nan,'k-','Linewidth',2);
    subplot(1,2,2);
    imagesc(nan);
    axis equal
    axis tight
    lx=length(x);
    b=15;
    if ~exist('ok','var') || length(ok)~=lx
       ok = false(lx,1);
    end
    if ~exist('ring','var') || length(ring)~=lx
       ring=false(lx,1);
    end
%     rx=cell(lx,1);
%     ry=rx;
    tx=cell(lx,1);
    ty=tx;
```

```
k=1;
while k <= lx
    yi=max(1,round(min(y{k}))-b):min(sb(1),round(max(y{k}))+b);
    xi=max(1,round(min(x{k}))-b):min(sb(1),round(max(x{k}))+b);
    set(border,'XData',[xi(1) xi(end) xi(end) xi(1) xi(1)],...
        'YData',[yi(1) yi(1) yi(end) yi(end) yi(1)]);
    subplot(1,2,2);
    imagesc(xi,yi,bild(yi,xi));
    axis equal
    axis tight
    hold on;
    [X,Y] = meshgrid(xi,yi);
    % [rx{k}, ry{k}]=refinepos(bild(yi,xi), x{k}, y{k},X, Y,step);
   if (~isequal(x{k}(1),x{k}(end)) || ~isequal(y{k}(1),y{k}(end))) &&...
           sqrt((x{k}(1)-x{k}(end)).^2+(y{k}(1)-y{k}(end)).^2) < 2*step;
       ring(k) = true; %#ok<AGROW>
       x{k}=[x{k};x{k}(1)];
       y{k}=[y{k};y{k}(1)];
    end
    [tx{k},ty{k}]=trackit(bild(yi,xi), x{k}, y{k},X,Y,step);
    %[tx{k},ty{k}]=interpsel(x{k}, y{k}, step);
    if ring(k)
       tx{k}=[tx{k} tx{k}(1)]; %#ok<AGROW>
       ty{k}=[ty{k} ty{k}(1)]; %#ok<AGROW>
    end
    %h=plot(x{k}, y{k},'*',tx{k},ty{k},'r',rx{k},ry{k},'+');
    h=plot(x{k}, y{k},'*',tx{k},ty{k},'r');
    set(h(2),'LineWidth',2);
```

```
    if ring(k)
       set(h(2),'Color','k');
    end
    hold off;
    title(sprintf('Curve number: %4i/%4i, Accepted: %d', k,lx,ok(k)));
    a=axis;
    btn=0;
    while btn ~= 1
       [gx,gy,btn]=ginput(1);
       if gx>a(1) && gx < a(2) && gy > a(3) && gy < a(4)
          ok(k)=true; %#ok<AGROW>
          if btn==2 && ~ring(k)
             x{k}=[x{k};x{k}(1)];
             y{k}=[y{k};y{k}(1)];
             tx{k}=[tx{k} tx{k}(1)]; %#ok<AGROW>
             ty{k}=[ty{k} ty{k}(1)]; %#ok<AGROW>
             ring(k) = true; %#ok<AGROW>
%               disp(sprintf('File: %15s, Curve nr.: %4i: DNA ring closed',...
%                  files(l).name, k));
          elseif btn==3 && ring(k)
             x{k}=x{k}(1:end-1);
             y{k}=y{k}(1:end-1);
             tx{k}=tx{k}(1:end-1);
             ty{k}=ty{k}(1:end-1);
             ring(k)=false; %#ok<AGROW>
%               disp(sprintf('File: %15s, Curve nr.: %4i: DNA ring opened',...
%                  files(l).name,k));
          end
```

```
else
   if btn==1
      ok(k) = false; %#ok<AGROW>
      break;
   elseif btn == 2
      input = inputdlg({'Length of orthogonal line',...
         'Number of iterative loops','Method'},...
         'Tracking Parameter',...
         1,{'4','1','1'});
      if ~isempty(input)
         if strcmpi(input{3}(1),'0')
            [tx{k},ty{k}]=interpsel(x{k}, y{k}, step);
         else
            nloops = round(str2double(input{2}));
            if ~isfinite(nloops) || nloops < 1, nloops = 1;end
            normlen = str2double(input{1});
            if ~isfinite(normlen) || normlen < 1, normlen = 1;end
            [tx{k},ty{k}]=trackit(bild(yi,xi), x{k}, ...
               y{k},X,Y,step,nloops,normlen);
         end
         if ring(k)
            if ~isequal(tx{k}(1),tx{k}(end)),
               tx{k}=[tx{k},tx{k}(1)];
            end
            if ~isequal(ty{k}(1),ty{k}(end)),
               ty{k}=[ty{k},ty{k}(1)];
            end
            if ~isequal(x{k}(1),x{k}(end))
```

```
                x{k}=[x{k};x{k}(1)];
            end
            if ~isequal(y{k}(1),y{k}(end))
                y{k}=[y{k};y{k}(1)];
            end
        end
    end
elseif btn==3
    input = inputdlg('Goto Curve');
    if ~isempty(input)
        input = str2double(input{1});
        if isfinite(input)
            k = input - 1;
            if k < 0, k = 0;end
            if k > lx - 1, k = lx - 1;end
        end
        break;
    end
end
end
if ring(k)
    set(h(2), 'Color', 'k');
else
    set(h(2), 'Color','r');
end
set(h(2),'XData',tx{k},'YData',ty{k});
end
k = k + 1;
```

```
    end
    save(files(l).name,'-append','x','y','tx','ty','ok','ring',...
        'nmperpx','step');
    clear ok ring nmperpx
end
```

trackdnasel.m:

```
function trackdnasel
[file,pfad]=uigetfile('*.*','Select Image to analyze');
if isequal(file,0)
    return
end
file=fullfile(pfad,file);
[sav,pfad]=uiputfile('*.mat','Save the results to...');
if isequal(sav,0)
    return
end
sav=fullfile(pfad,sav);
if strcmpi(file(end-2:end),'TXT')
    bild=load(file);
    answ=inputdlg(sprintf('Scan Size of the image in the file %s',...
        file));
    if isempty(answ)
        return
    end
    answ=answ{1};
    bheader{1}.length_x=str2double(answ);
    if ~isfinite(bheader{1}.length_x)
        [tmp,bheader]=stpread(answ);
```

```
        if ~exist('tmp','var') || isempty(tmp)
            return
        end
    end
elseif strcmpi(file(end-2:end),'STP')
    warning off
    [out,bheader]=stpread(file);
    warning on
    bild=flatrow(out{1}, 2);
else
    bild=imread(file);
    if size(bild, 3) > 1
        bild = rgb2gray(bild);
    end
    bild = double(bild);
    answ=inputdlg(sprintf('Scan Size in Angstrom of the image in the file %s',...
        file));
    if isempty(answ)
        return
    end
    answ=answ{1};
    bheader{1}.length_x=str2double(answ);
end
sb=size(bild);
nmperpx=bheader{1}.length_x/(10*sb(2));
imagesc(bild);
hold on;
```

```
M=median(bild(:));
S=mad(bild(:));
caxis(M+[-4 4]*S);
k=0;
templ=cell(1000,1);
x=templ;
y=templ;
while 1
   [xt, yt] = getpts;
   if length(xt)==1
      break;
   end
   k=k+1;
   if k>length(x)
      x=cat(1,x,templ);
      y=cat(1,y,templ);
   end
   save(sav, 'bild', 'x', 'y');
   x{k}=xt;
   y{k}=yt;
   line(x{k}, y{k});
end
close gcf;
x=x(1:k);
y=y(1:k);
save(sav, 'bild', 'x', 'y', 'nmperpx');
```

trackit:

```
function [tx,ty]=trackit(bild, x, y, X, Y, step, loopn, normlen)
if nargin < 7
   loopn = 1;
end
if nargin < 8
   normlen = 4;
end
if nargin < 6
   step = 2.5;
end
t = linspace(-normlen, normlen, 30);
% t = 30*pi/180*t;
% Estimate the number of points
d = sum(sqrt((x(2:end)-x(1:end-1)).^2 + (y(2:end)-y(1:end-1)).^2))/step;
tx = nan(1,floor(3*d));
ty = tx;
lt=length(tx);
ind = 0;
k=1;
%for k=1:length(x)-1
   % Direction to the next clicked point
   dx=x(k+1)-x(k);
   dy=y(k+1)-y(k);
   ind = ind + 1;
   tx(ind) = x(k);
   ty(ind) = y(k);
   while ind < 9*d
```

```
count = 0;
% Normalize to unit length
r=sqrt(dx^2+dy^2);
dx=dx/r;
dy=dy/r;
% First estimate for the next point
nx = tx(ind) + dx * step;
ny = ty(ind) + dy * step;
while (count < loopn)
   % Calculate the perpendicular line
   normx = nx - t * dy * step;
   normy = ny + t * dx * step;
%        phi0 = atan2(dx, dy);
%        normx = nx + step * sin(t+phi0);
%        normy = ny + step * cos(t+phi0);
   % Caclulate the image along the above line
   A = interp2 (X, Y, bild, normx, normy, '*linear');
   % Find the maximum, use a robust estimation:
   % find the biggest part which is over min(A) + range(A)/3
   % find thru these data a 2nd order polynomial, use the maximum
   % of this function as maximum estimate for the DNA
   [mxa,mxi]=max(A);
   mna=min(A);aind = A > mna+(mxa-mna)/3;
   fi = find(diff([0,aind,0]));
   [mxfi,fii]=max(diff(fi));
   aind=fi(fii):fi(fii+1)-1;
   if length(aind) > 5
      p=polyfit(t(aind), A(aind),2);
```

```
            tmax=-p(2)/(2*p(1));
            if tmax < t(1) || tmax > t(end)
               tmax=t(mxi);
            end
         else
            tmax = t(mxi);
         end
            nx = nx - tmax * dy * step;
         ny = ny + tmax * dx * step;
%           nx = nx + step * sin(phi0 + tmax);
%           ny = ny + step * cos(phi0 + tmax);
%
%           sa = sum(A);
%
%           % center of mass
%           nx = sum (normx .* A)/sa;
%           ny = sum (normy .* A)/sa;
         dx = nx - tx(ind);
         dy = ny - ty(ind);
         r = sqrt(dx^2 + dy^2);
           dx = dx / r;
            dy = dy / r;
             nx = tx(ind) + dx * step;
         ny = ty(ind) + dy * step;
           count = count + 1;
      end
   ind = ind + 1;
```

```
        if ind > lt
            tx=[tx nan(1,lt)]; %#ok<AGROW>
            ty=[ty nan(1,lt)]; %#ok<AGROW>
            lt=length(t);
        end
         tx(ind) = nx;
        ty(ind) = ny;
%         hold on;
%         plot(tx, ty,'x-');
%         pause;
        if ind > 2 && sqrt((tx(ind)-x(end)).^2 + (ty(ind)-y(end)).^2) < step
            break
        end
    end
% end
 tx=tx(isfinite(tx));
ty=ty(isfinite(ty));
```

APPENDIX B

MATLAB CODE FOR ANALYZING GELS

```
MaxSites=17;
MaxProt=0;
Iterations=40000;%stable statistics need 40,000
nbins=40;
B=[];
y=[];
n=[];
Cumulative=[];
for k=1:Iterations
    AvaCutSites2=AvaCutSites;
    A=randperm(MaxSites);
    for p=1:MaxProt
        B(p)=A(p)+1;
    end
    AvaCutSites2(B)=[];
    for q=1:(MaxSites-MaxProt+1)
        y(q)=AvaCutSites2(q+1)-AvaCutSites2(q);
    end
    Cumulative=[Cumulative y];
end
AvaCutSites2; %diagnostic
Cumulative; %diganostic
[n,xout]=hist(Cumulative,nbins);%Histogrammed Y.X data
n2=n.*xout; %Gel intensity is proportional to length
%bar(xout,n)
bar(xout,n2)
OutPut=[xout' n' n2'];
```

```
csvwrite('0_sites.csv',OutPut);
```