Insight into the impact of methylation and demethylation pathways in DNA physical properties

Modesto Orozco's lab

Pen-name: Biomed

Insight into the impact of methylation and demethylation pathways in DNA physical properties

ABSTRACT

Methylated DNA has recently been a novel object of study for its implications in biological pathways through the regulation of the gene expression. Otherwise, its acting method and characteristics still remain unknown. In this study we unveil the physical properties of DNA during methylation and demethylation pathways (Met. - Hydroxymet. - Formylmet. - Carboxy. and Removal of methyl group) of the CpG cytosine to obtain a proper idea on how these modifications affect DNA stability. We observe that cytosine methylation heavily decreases DNA flexibility in both circularization assay and Monte-Carlo based simulations. We also describe that both methylated and hydroxymethilated DNA have got an increased chance of getting pieces with high bending, at around 15-25 degrees, due to high stiffness and high local curvature values. Overall, we have shown that both methylation and demethylation have got direct implications in DNA flexibility and, as a result, in the epigenetic regulatory system.

Keywords: *methylation; demethylation; DNA physical properties; DNA flexibility; Monte Carlo; DNA bending; epigenetics; cancer*

Influència de les vies de metilació i desmetilació en les propietats físiques de l'ADN

RESUM

Les metilacions en l'ADN han estat recentment descrites com a un nou objecte d'estudi degut a les seves implicacions en processos metabòlics d'entre els quals s'inclou la regulació de l'expressió gènica. No obstant, el seu mètode d'actuació i les característiques que confereix a la molècula d'ADN encara no s'han identificat. En aquest estudi desvetllem les propietats físiques de l'ADN durant els diferents processos de metilació i desmetilació (Met. – Hidroximet. – Formilmet. – Carboxy i eliminació del grup metil) de les citosines localitzades al CpG; això ens permet obtenir una idea de com aquestes modificacions afecten a la seva estabilitat. Observem que la metilació redueix dràsticament la flexibilitat de l'ADN tant en l'assaig de circularització com en les simulacions fetes emprant l'algoritme de Monte-Carlo. També descrivim que l'ADN metilat i hidroximetilat té més probabilitats de donar lloc a fragments amb un grau de curvatura més elevat, al voltant dels 15-25º, a causa d'alts valors en la rigidesa i en la curvatura local. En general, hem mostrat que tant la metilació com la desmetilació tenen implicacions directes en la flexibilitat de l'ADN i, per tant, en la regulació de l'expressió gènica.

Paraules clau: *metilació; desmetilació; propietats físiques de l'ADN; flexibilitat de l'ADN; Monte-Carlo; curvatura de l'ADN; epigenètica; càncer*

Influencia de las vías de metilación y desmetilación en las propiedades físicas del ADN

RESUMEN

Las metilaciones del ADN han sido recientemente descritas como un nuevo objecto de estudio debido a sus implicacions en procesos metabólicos de entre los cuales se incluye la regulación de la expresión génica. No obstante, su método de actuación y las características que confiere a la molécula de ADN siguen aún sin haberse identificado. En este estudio nos adentramos en las características físicas del ADN durante los diferentes procesos de metilación y desmetilación (Met. – Hidroximet. – Formilmet. – Carboxi y la eliminació del grupo metilo) de las citosinas localitzadas en los motivos CpG; esto nos permite obtener una idea de cómo estas modificaciones pueden afectar a la estabilidad del ADN. Observamos que la metilación reduce drásticamente la flexibilidad del ADN tanto en los ensayos de circularización como en las simulaciones hechas haciendo usando del algoritmo de Monte-Carlo. También describimos que el ADN metilado y el hidroximetilado tienen más probabilidades de dar lugar a fragmentos con un grado de curvatura más elevado, sobre los 15-25º, a causa de altos valores en la rigidez y en la curvatura local. En general, hemos mostrado que tanto la metilación como la desmetilación tienen implicaciones directas sobre la flexibilidad del ADN y la regulación de la expresión génica.

Palabras clave: *metilación; desmetilación; propiedades físicas del ADN; flexibilidad del ADN; Monte-Carlo; curvatura del ADN; epigenética; cáncer*

INDEX

0. Introduction

The Research Work is a perfect chance to develop a hypothesis and work on a field for a period of time to increase your knowledge in order to obtain a final conclusion. This project is focused on the simulation of 3D structures of naked B-DNA and the understanding of the algorithms behind the computational models. Furthermore, experiments and simulations on higher order DNA structures like chromatin are performed because knowing the 3D structure of chromatin allows us to achieve quicker transcription profiles and determine how epigenetics affect chromatin stability. DNA modifications play a key role in the structure of chromatin (for example the formation of nucleosomes) and, as a result, the regulation of gene expression. Understanding how these modifications change DNA flexibility will help us understand their pathway on a biological and molecular level. We will focus on cytosine methylation at a CpG step as an epigenetic mark to study DNA flexibility and the different intermediates of the demethylation process, which involves a series of chemical reactions leading to the removal of the methyl group.

This cutting edge research includes understanding the biological role of DNA modifications, its effect on DNA flexibility and determining the three-dimensional structure of naked B-DNA at a kilo base-pair (kbp) level through simulations, which is still not known. During this work challenging problems are faced and discussed in order to solve them with the help of Jürgen Walther, a PhD student working in the Structural & Computational Biology unit of the IRB Barcelona (Institute for Research in Biomedicine) in the group of Molecular Modelling and Bioinformatics and Isabelle Brun-Heath, the Experimental Lab Director of the same group, who has helped with the experimental part of the project.

This work is divided in six main sub-topics. In the first one DNA is described as the molecule that carries all the genetic information and we take a close look at the structure and components of the nucleotides as well as the Chargaff's rule. The second one focuses on the theory behind the circularization assays as well as the characteristics of the 1st dimension and 2nd dimension electrophoresis gels and the Phosphoramidite method used to synthesise the DNA strands used in the study. During the third sub-topic we describe the theory behind our computational model whereas in the fourth one the procedure of the Monte Carlo based simulation and the results obtained are described in detail. In the fifth sub-

topic the procedure for the experimental part as well as the results are also described and finally in the sixth sub-topic a comparison between both parts is made in order to get a final conclusion.

1. General Background

The molecular conformation of DNA *in situ* has been a point of research since the first DNA conformation discovered by Watson and Crick of the right-handed double helix, known nowadays as the B-DNA structure. Since then, dozens of alternative double-stranded conformations have been identified including the A-DNA, a right-handed double stranded DNA conformation more compacted than Watson and Crick's B-DNA model and associated with dehydration cycles, the Z-DNA, a left-handed DNA structure that winds to the left instead of to the right and which is involved in high active gene expression regions, and Holliday junctions, four double-stranded DNA arms joined together that adopt different conformations depending on the buffer salt concentrations and the nucleotide sequences. Even triplex and quadruplex conformations have been recently described. [1, 2, 3] The lack of techniques to study their *in situ* conformation has resulted in the study of only a few of them, regarding special interest in A-DNA and B-DNA, which have shown key biological significance. In the following section the structure of B-DNA, the most common conformation of DNA in cells, will be described in detail.

1.1 DNA

Deoxyribonucleic acid, commonly known as DNA, is the life molecule which carries all the genetic information necessary for cell development, homeostasis, growth and reproduction. This information can be expressed by genes, short DNA fragments which have got a certain function after transcription (long/short non-coding RNA) or after translation (proteins). [3, 4]

DNA is made of organic molecules called nucleotides, which are composed of a five-carbon sugar (deoxyribose), a phosphate group and a nitrogenous base. The nitrogenous base can be either a thymine (T), a cytosine (C), a guanine (G), or an adenine (A). [4, 5] The phosphate group plays a key role in the formation of the phosphodiester bond that links two consecutive bases (Figure 1).

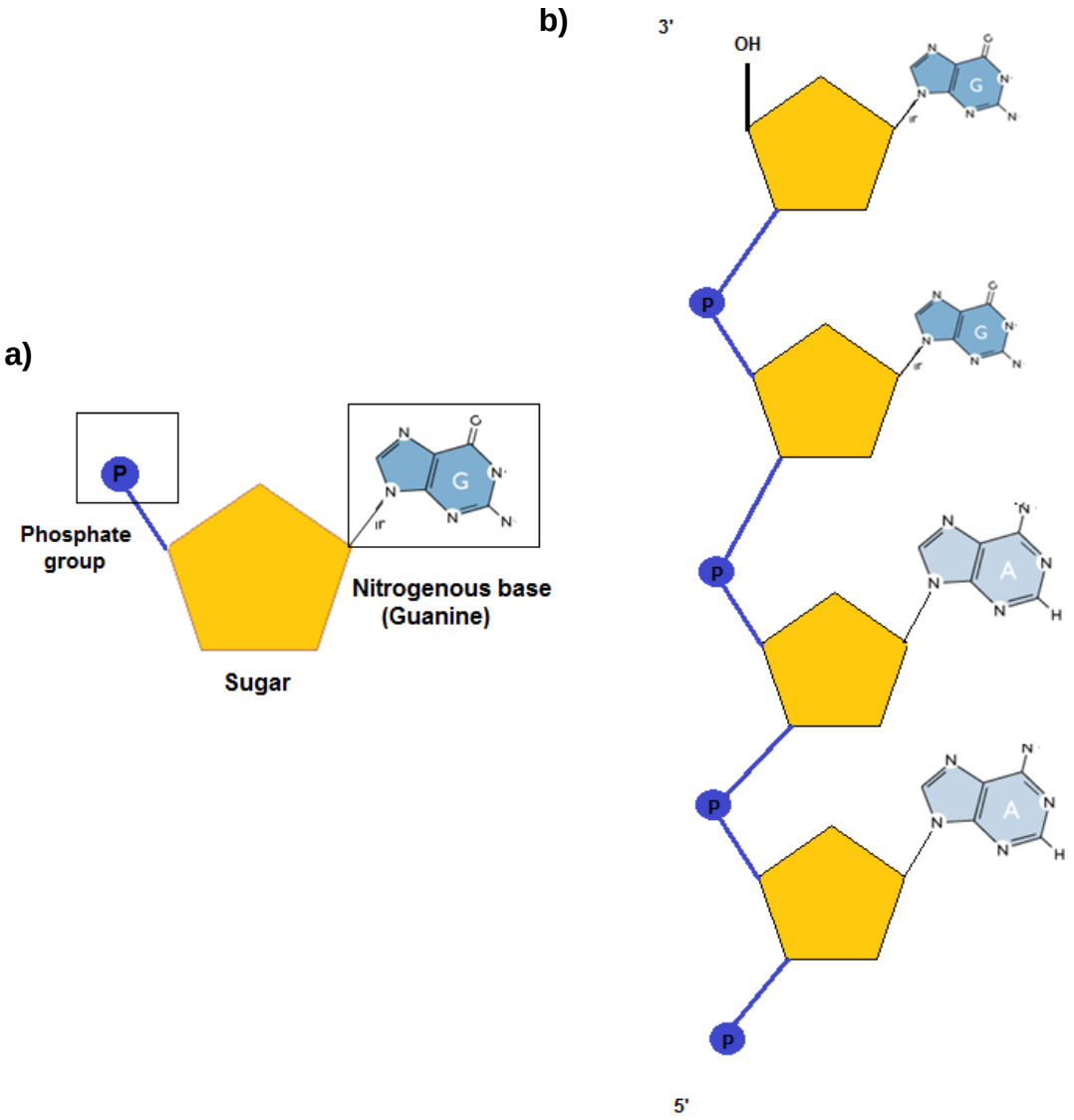

Figure 1. Nucleotide schematization. **a**) Nucleotide components: the sugar, which is the structural basis of the nucleotide, the nitrogenous base, which bonds nucleotides of two different DNA strands by the formation of H-bonds between them, and the phosphate group, which acts as a linker between two consecutive nucleotides. **b**) Representation of the linking action of the phosphate group, which creates a phosphodiester bond between nucleotides and attaches them. (Own authorship)

1.2 Chargaff's rule

Deoxyribonucleotides can be divided in pyrimidines (C-T) and purines (A-G). Both are aromatic compounds that differ from each other because purines are fused to an imidazole ring. [6]

The Chargaff's rule states that a pyrimidine is always complementary to a purine, and that the total amount of guanine is equal to the amount of cytosine and so the same ratio happens between adenine and thymine, without taking into account mutations. Bases interact through hydrogen bonds (H-bond) between nitrogens and hydrogens and between oxygens and hydrogens respectively. Two H-bonds can be formed between A and T whereas three can be formed between C and G due to the chemical structure of the bases (Figure 2). [2, 6, 7]

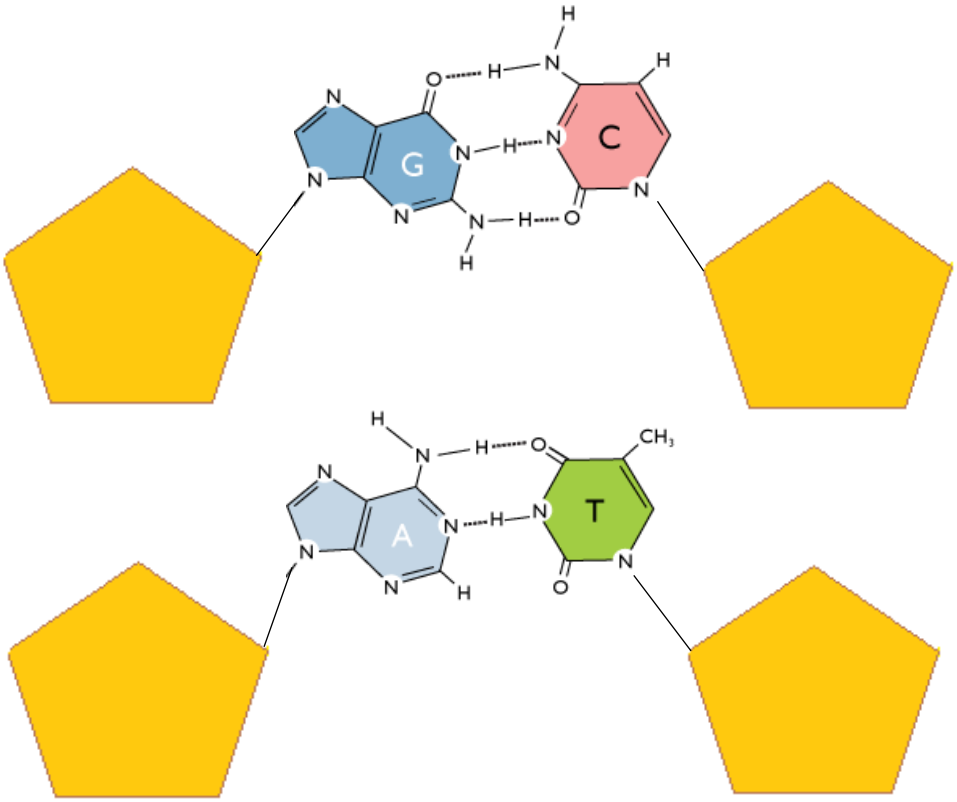

Figure 2. Representation of the Chargaff's rule. Pyrimidines (C-T) are always complementary with purines (G-A). Three hydrogen bonds are formed between C-G due to having two highly electronegative nitrogens (N) and one electronegative oxygen (O) in each base, which can bind together with a linker, the hydrogen (H), which is extremely electropositive. Between A-T there can only be formed two hydrogen bonds as the adenine has got a hydrogen instead of a nitrogen, which denies the chemical formation of another Hydrogen bond with the O, resulting in only two H-bonds attaching A-T bases. (Own authorship)

1.3 B-DNA

B-DNA is a DNA structure first seen by Rosalin Franklin and later on described and proposed by Watson and Crick in 1953. It's a right-handed double stranded DNA conformation with about 10-10,5 base pairs per turn. In this model, bases have to fit between them. Iy means that a pyrimidine (C-T) is always paired with

a purine (C-G) located in the other strand, by following the Chargaff's rule. A single strand of DNA contains a 3'-end with a free –OH group at the first nucleotide and a 5'-end at the last nucleotide which frequently contains a phosphate group in order to bind with other DNA molecules. [8, 9]

The B-DNA structure is made of two antiparallel coiled DNA strands, which results in the formation of the double helix of DNA. [8] It means that if one strand goes from 3' to 5', the other one will be from 5' to 3' just because replication happens in the way that DNA polymerase reads from 3' to 5' and synthesises from 5' to 3', leading to an antiparallel DNA fragment which will coil altogether with the other strand due to the formation of H-bonds between complementary nucleotides (Figure 3). [8, 9, 10]

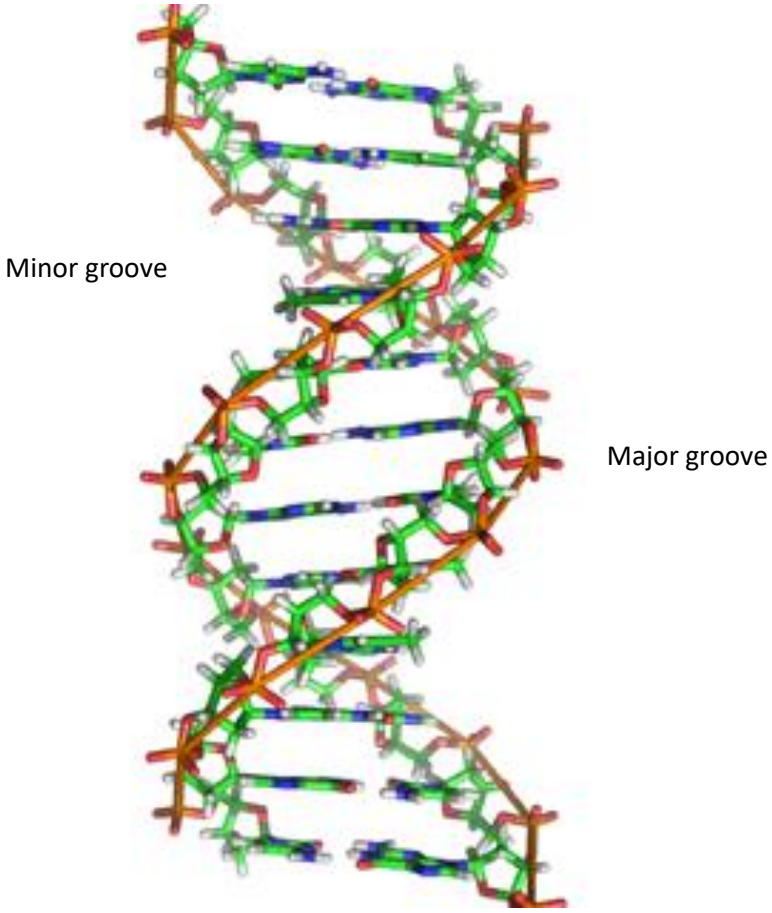

Figure 3. Schematic representation of the B-DNA model. [10] It's shown the right-handed double antiparallel DNA strand with major and minor groove pointed out in the figure, which represent the zones with maximum and minimum distances between both DNA strands.

2. Theoretical background for the experimental part

2.1 Chromatin

In eukaryotes, DNA is usually found forming chromatin, a level of organization in which DNA, which is negatively charged due to the phosphate group, wraps around specific positively charged proteins, called histones. The process leads to the formation of DNA-protein cores, nucleosomes. This high level of organization constraints a wide range of processes involving DNA, such as replication, transcription, repairing or even meiotic recombination by regulating the expression of the genes implicated in those processes. [9, 11, 12] The nucleosome core is composed of histone octamer units typically enfolded by 147 bp of DNA (Figure 4). [11]

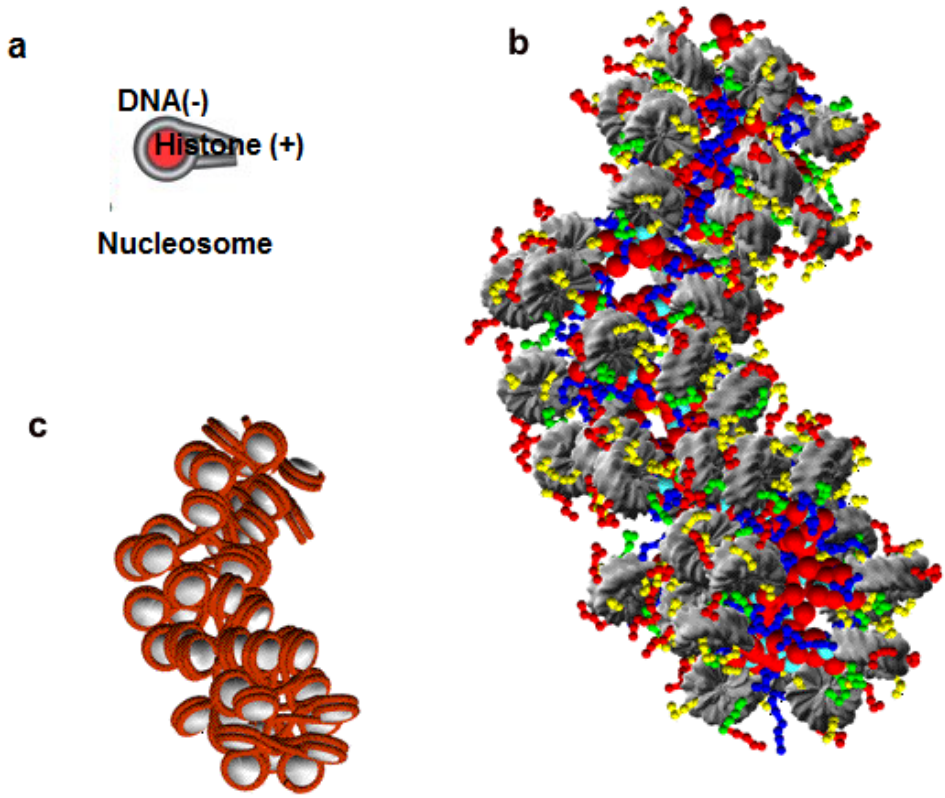

Figure 4. Representation of chromatin. [11] **a**) A nucleosome, which is the basis of a chromatin fibre, made of positively charged histones and negatively charged DNA wrapped around. **b**) Representative oligonucleosome simulation. **c**) Simple chromatin representation. It's seen that it's made of lots of nucleosomes folded around them and separated by linker DNA sequences.

As a conclusion, chromatin provides a dynamic platform that controls all processes involving DNA within the nucleus. There are four mechanisms that modulate chromatin compaction: ATP-dependent complexes which use the energy of ATP to slide the nucleosomes and exchange histones from the chromatin fibre, [11, 12, 13] histone variants that have different properties than the canonical ones and provide specific domains in the chromatin fibre, [14, 15] histone chaperones that regulate histone exchange during transcription and finally post-translational modifications of the histones and DNA modifications. [16]

2.2 DNA modifications and their role

Chromatin structure never remains the same, it's constantly dynamic depending on the requirements of the cell and the epigenetic marks through the environment. This chromatin structure changes as post-translational modifications are applied to the histones or modifications are applied directly to DNA. Histones are composed of a globular domain that mediates its interaction with the octamer and an amino-terminal of 20-35 residues (called 'tails') that does not contribute to the structure of individual nucleosomes, but can carry post-translational modifications. The histone tails play a key role in the folding of nucleosomes with DNA and condensation.[11-16]

DNA can also suffer modifications which will result in an increase or decrease of its physical properties, including flexibility and, as a result, a change in the probability of forming a nucleosome due to being more flexible or more rigid. DNA can only suffer methylations due to chemical and physical incompatibilities with other groups. The effect of methylated DNA on biological processes remains largely unknown, so this project will help us understand both methylation and demethylation processes on DNA, DNA physical properties and the role within chromatin formation and gene expression.

2.2.1 DNA Methylation and demethylation

Methylation is a covalent epigenetic mark that happens with the addition of a methyl group to a lysine or an arginine of a protein, or to a cytosine in the DNA. Histone methylation plays a key role in biological processes including transcription, single transduction, development, and cellular proliferation and differentiation.[19-21] Histone methylation is associated with either transcriptional

repression or activation, but it's not clear the biological impact of DNA methylation and demethylation.[22] In this project we are studying the methylation of the carbon 5' of the cytosine ring, leading to 5-methylcytosine (5-mC) which has been proposed to disperse chromatin and decrease nucleosome formation, being associated with gene repression. DNA methylation is performed by DNA methyltransferases (DNMTs), which include a series of metabolic pathways object of this study. [21, 22, 23, 24]

Cytosine methylation occurs at a CpG site, which conforms a cytosine nucleotide linked with a guanine by a phosphate. About an 80% of the CpG domains in genomic DNA are methylated, but there are some exceptions such as embryonic stem cells, where there are cytosines methylated in non-CpG regions or even in germ-line tissues, where CpG islands (clusters) remain unmethylated (active transcription regions) near the promoters. On the other hand, DNA demethylation involves a series of metabolic pathways resulting in the complete removal of methyl groups and seems to be necessary for gene expression. DNA demethylation can be either active or passive. The active one refers to an enzymatic pathway that removes directly the methyl group from 5-mC through different processes carried by the ten-eleven translocation (TET) family enzymes. These enzymes bind to the CpG domain and prevent the action of new DNMTs; but they play a secondary role by converting 5-mC to 5-hmC with the addition of a hydroxyl group, then from 5-hmC to 5-fC by converting the hydroxyl intro a formyl group, from 5-fC to 5-caC with the modification and insertion of a carboxyl group and finally this group is removed from the cytosine through the hydroxylase activity carried by a decarboxylase, leading to a cytosine that can be methylated again. Passive DNA demethylation refers to the loss of 5-mC during successive rounds of replication in the absence of functional DNA methylation maintenance machinery (Figure 5). [19-25]

It's shown that this active way to remove the methyl-group causes changes on DNA structure. Demethylation sometimes has got a biological role when required by the cell such as rich promoter zones of 5-fC during embryonic stages, which are associated with highly active regions. [25, 26] But the remaining question is, why removal of the methyl group can't be performed on a single step? Which advantages, in case there are, does it confer to the cell or to a possible increase

of gene transcription? Is DNA changing flexibility between each of these methylation and demethylation paths? In this study we are comparing in both simulations and experimental assays DNA flexibility along the different steps of methylation and demethylation. In the following section details of the experimental setup are described.

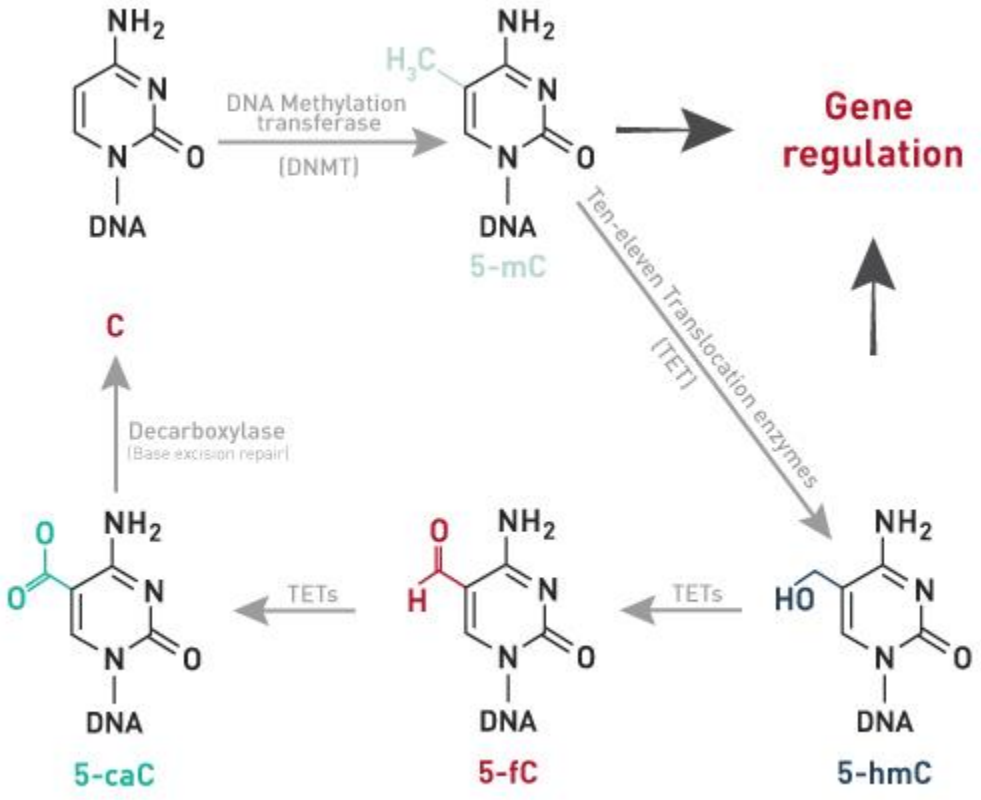

Figure 5. Dynamics of the passive demethylation pathway. [25] During this process, DNA methyltransferase methylates the cytosine by adding a methyl group, which affects gene expression. Later on, different TET enzymes play a key role in demethylation by adding first a hydroxyl group, forming 5-hmC. Then it's converted to 5-fC by changing the hydroxyl to a formyl group and finally the formyl group is changed again for a carbon which can be removed by a decarboxylase, leading to the complete removal of the initial methyl group.

2.3 The Phosphoramidite Method for DNA synthesis

The phosphoramidite method was pioneered by Marvin Caruthers in the early 1980s and right now it has been firmly established as the method of choice for DNA synthesis. It consists of a series of steps that include detritylation, activation and coupling, capping, oxidation and deprotection in order to synthesize DNA. [27]

At the beginning of the process, the oligonucleotides are attached to the resin with a 5'-DMT protecting group (DMT = 4,4'-dimethoxytrityl) that prevents the polymerization during resin functionalization, but it has to be removed in the detritylation so the oligonucleotide synthesis can proceed. The mechanism of detritylation consists in the addition of an acid which emits H^+ particles that later on will react with the DMT protecting group and, as a result this group is removed and we obtain a free 5' carbon, which lately will react with the following nucleoside thanks to the previous addition of a phosphate group and the formation of a phosphodiester bond (Figure 6). [27, 28]

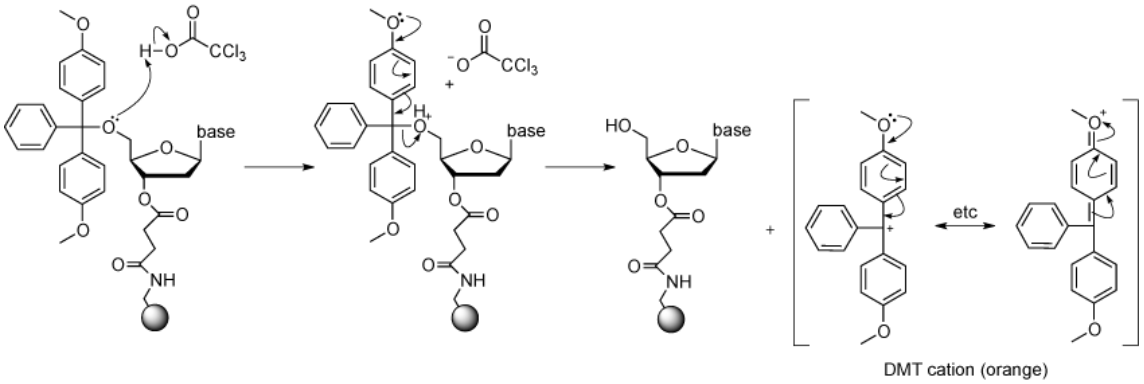

Figure 6. Schematic representation of the detritylation process. [27] An acid is applied in order to remove the DMT protecting group by making it react with H^+ particles. The 5' carbon gets free as a result.

In the next step, called activation and coupling, the attached nucleoside, which has lost the DMT group, is ready to react with the next base and form a phosphodiester bond. To do that, an excess of the appropriate nucleoside is mixed up with an activator, which is tetrazolium or a derivate. The diisopopulamino group is protonated by the acting of the activator and rapidly the 5'-OH group of the support nucleoside is attacked by the phosphorus protonated atom, the phosphorus-oxygen bond is created and as a result phosphite trimester is formed (Figure 7). [27, 29, 30]

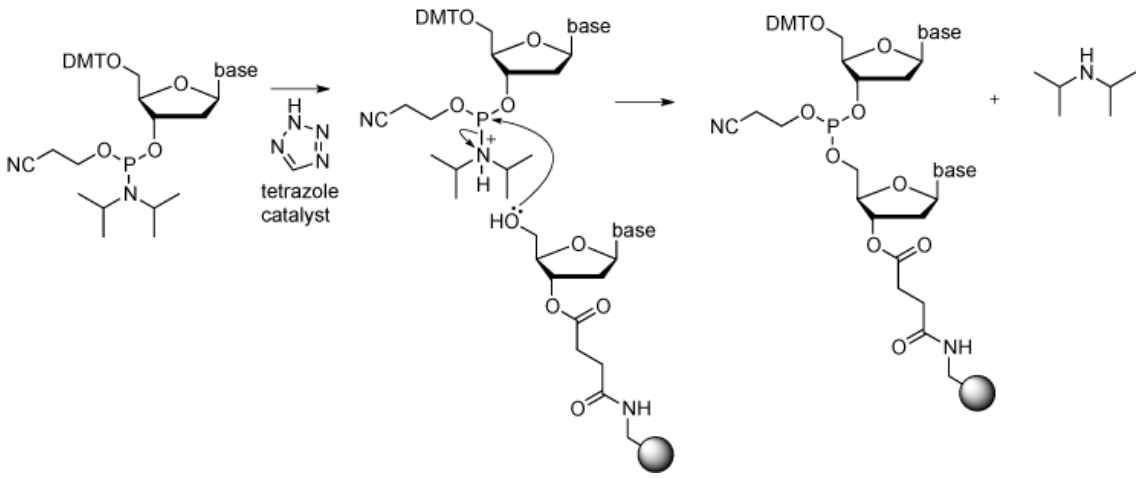

Figure 7. Schematic representation of the activation and coupling process. [27] Tetrazole is a catalyst that increases the formation of the phosphodiester bond. Rapidly the 5'-OH group of the support nucleoside is attacked by the phosphorus atom, leading to the formation of the phosphite triester.

The next step is the capping, which is no more than a test to avoid possible mutations due to unreacted 5'-OH groups on the resin-bound nucleotide that can lead to further deletions or mutations to the nucleotide chain. Two capping solutions are introduced in the DNA synthesizer: acetic anhydride and N-methylimidazole (NMI). These two chemical products prevent possible detritylation of the nucleoside by the acting of acetic acid. The pyridine ensures the maintenance of a basic pH (Figure 8). [27, 31]

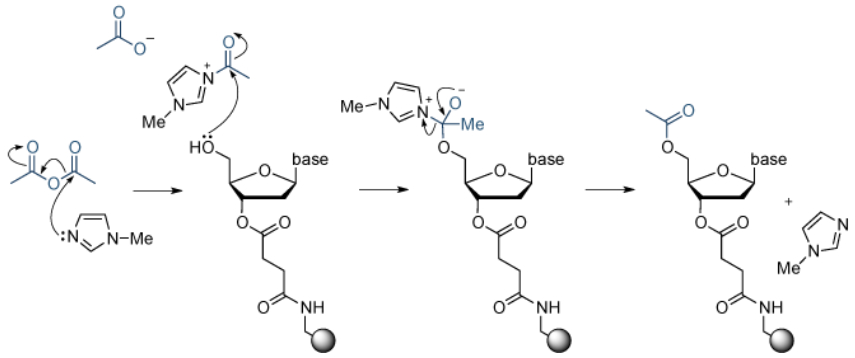

Figure 8. Representation of the capping process. [27] This transition step has the only function to avoid possible mutants or mismatches during the previous process. Two capping solutions are introduced which act on the 5'-OH of pyridines and purines by maintaining their pH levels, so they can't be detrytilated.

Everything is followed by an oxidation step, just because the phosphite-triester that was formed in the coupling is unstable to the reactant of an acid and must be changed prior to the following steps. This process is achieved by iodine oxidation, in which an oxygen double bond is added to the phosphite trimester, being more stable than the previous chemical configuration. A 2-cyanoethyl group is also added in order to protect the phosphorus from undesirable foreign reactions which are not needed (Figure 9). [27, 29, 30]

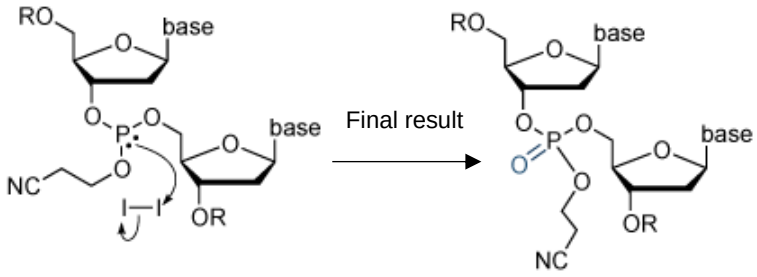

Figure 9. Schematic representation of the oxidation step. [27] This process involves the iodine oxidation of the phosphite-triester, which leads to a double bond oxygen formation, which is more stable. A 2-cyanoethyl group is also added in order to protect the phosphorus from foreign reactions.

After coupling, capping and oxidation, the DMT protecting group at the 5'-end of the resin bound DNA chain has to be removed in order that the next –OH group can react with it again. Deprotection is usually done with trichloroacetic acid as it's a quite quick and effective reaction; as a result, an orange colour is produced due to the cleaved DMT carbocation. After that, this cycle is repeated as many times as necessary with the DNA synthesiser, once for each base, to produce the final DNA single strand sequence (Figure 10).

We have to point out that cytosine nucleotides have already been bought with the desired modifications for the experiment done in our study, so it was only necessary to link them in the chain following the previous reported procedure. [27, 29, 30]

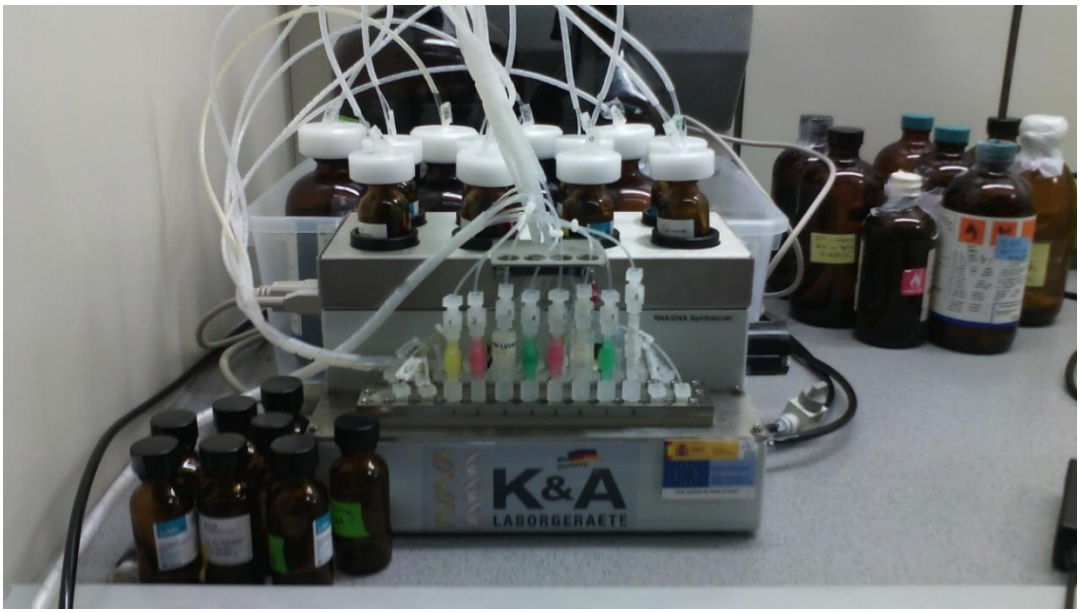

Figure 10. DNA synthesiser. This machine is used to synthesise DNA chains. Each of the tubes contain a different ribose. Each nucleoside is added following the Phosphoramidite Method. This procedure is repeated as many times as needed by introducing a command to the computer that controls the synthesiser. (Own authorship)

After having our DNA molecule synthesised, to define the methylation and its role on DNA flexibility, circularization assays are done in order to analyse how many DNA molecules are needed to circularize, which is directly related with its flexibility, so a molecule that tends to form circles means that it has a high flexibility. A close look at its principles is taken in the following point.

2.4 Theory behind circularization assays

The circularization assay is a test used to determine the flexibility of a DNA fragment within its capability to circularize. If it's more flexible, DNA will tend to wrap around other proteins, like histones, and form nucleosomes, whereas if DNA is less flexible, it's going to have more difficulties to circularize and less nucleosomes will be formed, increasing gene expression. [32, 33, 34]

Circularization assays always have the same principle: first of all the phosphorylation, which consists in the addition of a phosphate group to the 5'-end of DNA, is done so later on a phosphodiester bond can be formed with the –OH group of the 3'-end of another DNA fragment. The next step is the annealing, in which both DNA oligos from forward and reverse are cooled down from an

initial temperature of around 92ºC so DNA complementary strands will start forming hydrogen bonds between them (due to the cool down) until the double stranded structure is achieved. The most important step begins now and it is the ligation, in which an enzyme called DNA ligase is used to connect the phosphate group located in the 5'-end of one fragment with the –OH group of the 3'-end of another fragment. During this process, DNA fragments are attached between them until they are flexible enough to circularize. Finally, an ethanol precipitation is done to purify DNA and extract ligase residues and other pollutants from the samples. In this process sodium acetate (NaOAc) is used because as Na^+ is charged positively it will bind to the negatively charged DNA and will make it less hydrophilic and therefore less soluble, leading to DNA precipitation and the formation of a white surface in the Eppendorf tube, the pellet. Ethanol is necessary to favour the interaction between Na^+ and PO^{3-} (Figure 11).

Once precipitated, the DNA pellet is washed with ethanol 70% to remove the excess of salt, air dried and resuspended in water.

After circularization, an electrophoresis assay will provide us information from the circularization by allowing us to quantify circular versus linear DNA. Electrophoresis helps separating both linear and circularized molecules with different molecular weight and isoelectric points, which lately can be visualized by previously staining the gels with ethidium bromide and watching them in the dark chamber with UV rays.

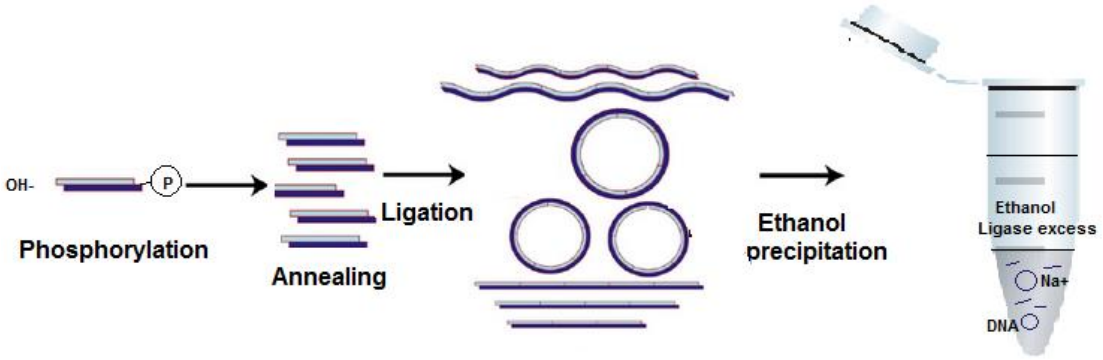

Figure 11. Representation of the circularization assay. After the single strand DNA is synthesised, the 5'-end is phosphorylated, so DNA can form phosphodiester bonds with other DNA molecules. Later on, annealing is done in order to obtain a double stranded DNA by decreasing the temperature, which increases the approximation of single strands and the formation of H-bonds between the complementary bases. The ligation is done with a DNA ligase to circularize our DNA, because the enzyme will attach DNA fragments until they are enough flexible to circularize. Finally the ethanol precipitation is done, which consists on a DNA purification method to remove ligase residues and other pollutants. (Own authorship)

2.5 Electrophoresis

The electrophoresis is a method used to separate and analyse macromolecules, such as DNA, RNA or proteins with different molecular weight. The gel can be composed of either polyacrylamide or agarose, which will produce tiny pores after polymerization through which our molecules will get across. The gel is then placed in an electric field that allows the migration of the sample toward the electrode of the opposite charge. In our case, as DNA is negative due to the phosphate group, it migrates towards the anode.

2.5.1 1st dimension electrophoresis

The 1st dimension electrophoresis consists on a 5% native PolyAcrylamide Gel Electrophoresis (PAGE) gel. It is run at a voltage of around 30 V but this can be modified depending on the sample and the time of migration. A molecular weight marker is also loaded on the gel to determine the size of our fragments. The tank of the apparatus is filled with running buffer, the optimal medium in which electricity goes through. [37, 38, 39]

The final result of the gel is a series of bands whose size can be determined thanks to the molecular-weight marker. To be able to visualise the bands, the gel

has to be stained with a fluorescent DNA intercalant like ethidium bromide and later on observed in a dark chamber with UV. [38, 39]

2.5.2 2nd dimension electrophoresis

The 2nd dimension electrophoresis is performed to separate the bands of the 1st dimension gel corresponding to the linear fragments from the ones of circular DNA. It consists of an 8% Native PAGE gel containing chloroquine. [41, 42] Chloroquine intercalates between the two strands of DNA and converts highly negatively supercoiled topoisomers into less supercoiled ones [43] allowing the separation of linear and circular DNA due to the fact that circular DNA is more coiled that linear DNA and once it gets uncoiled, it has different properties, which will allow them migrate in a different way according to their different isoelectric points. [40]

3. Theoretical background for simulation part

3.1 DNA helical parameters

DNA helical parameters describe the spatial location of the DNA bases respect to either an axis or another base; the sugar phosphate backbone is not taken into account when dealing with simulations involving helical parameters. The six inter base pairs parameters, which are rise, twist, shift, roll, tilt and slide determine the local conformation of the double helix strand by describing each base pair location from the consecutive ones. [44, 45, 46]

Twist, roll and tilt are the three rotational components. Positive rotation is defined by the right-hand rule in which if the rotation is produced towards the right side, the value will be positive and if it's towards the left side, a negative value will be attributed. Tilt and roll are both the rotational components for the x-y axis whereas twist is representative for the rotation around z axis. A negative roll closes the angle between two base pairs towards the major groove. On the other side, a positive tilt opens the angle of the base pairs towards the first strand. [46-48]

It's really important to take into account both positive and negative values of the parameters as well as the proper reference for the x, y, z axis for each base pair. Otherwise, wrong values will be obtained after doing calculations. [49]

The model with the six helical parameters describing the position of each base-pair assumes that the base-pairs are planar. However, the base pairs are not planar. As an example, the twist of an A-T base pair in B-DNA is ranged between 15º to 20º. [46] According to the fact that base pairs are not planar, the model obtained from the six inter base pair parameters will only provide us a rough model of the helix [46, 47] because we are not obtaining the most accurate results, as we are not looking at the intra base pair parameters between two complementary bases, which will describe us the angle between them. As our study focuses on the flexibility of the whole double helix of DNA, this rough error just plays a minor role as the inter base pairs parameters are enough for the study and analysis of the flexibility. [49, 50, 51] In the figure there are represented both inter base pair parameters (twist, roll, rise, tilt…) and intra base pair parameters (stagger, stretch, shear, buckle…) which form the complete list of helical coordinates for DNA (Figure 12).

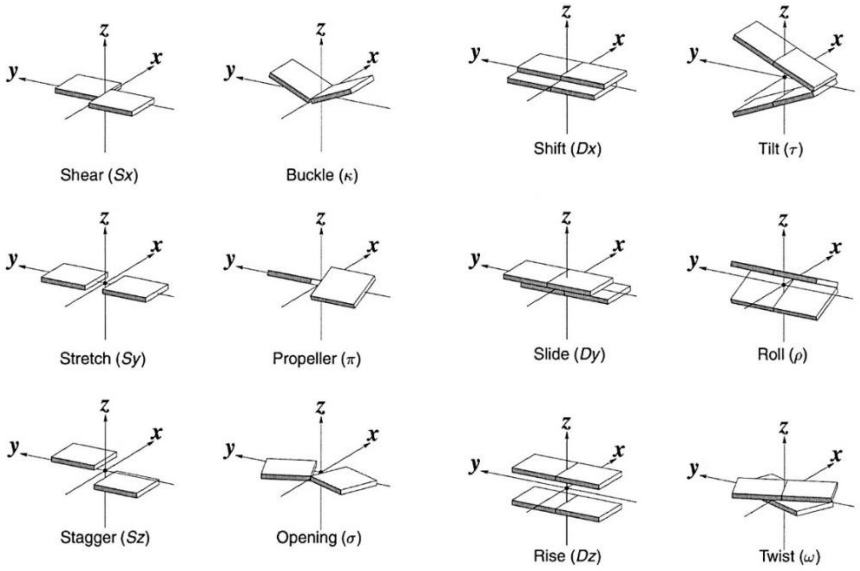

Figure 12. Representation of the inter base (roll, twist, rise, shift, slide, tilt) and intra base pair (stagger, stretch, shear, buckle, propeller, opening) parameters. [51] They provide information about the exact location of two different base pairs between them (inter base pair) or the location of two complementary bases (intra base pairs).

3.2 DNA flexibility and bending

Although DNA was first described as a rigid molecule, it was then described as a dynamic structure that can wrap around itself or other structures, like histones, and suffer further modifications. These physico-chemical properties of DNA directly depend on the nucleotide structure and the modifications, which affect their physical properties, like the flexibility or the bending. [50]

Bending flexibility is described as the inverse of the stiffness, because a high stiffness constant will result in a higher force between inter base pairs when moving the base-pairs apart and that would decrease the flexibility of the molecule. Furthermore, low stiffness levels will result in an increase of the flexibility as a result of lowering the force produced between those base pairs. As DNA is both polymorphic and heterogeneous in sequence, the stiffness of each base-pair step depends on its underlying sequence. [52-54]

Bending is known as the capability of DNA for being flexible. It is expressed in degrees (º) and it plays a key role in biological processes requiring DNA interaction with proteins. Some of them can be as important as some transcription factors which increase their interaction with promoter regions due to the bending of DNA. The base pairs play a pivotal role in this type of bending and so in DNA modifications. [55, 56]

DNA molecules often have a preferred bending direction which is determined by the stability of the bases located on the top of the strand. [55] If unstable bases are found in one strand of our DNA helix, the molecule is going to bend in the opposite site. The increase of the bending angle is also connected with an increase of the rotational DNA parameters, especially associated with high roll values in which the opened angle will go towards the minor groove. This is the effect studied and appreciated in biological processes like nucleosomes. [56, 57]

One possibility to compute the DNA bending is to trace the helical trajectory of DNA over the arc length of the whole sequence. Unfortunately, this method doesn’t take into account possible DNA-binding proteins attached that lead to nucleosome formation. [55, 57]

3.3 Hooke's law and bending stiffness.

In our simulation DNA is represented as planar base pairs connected each other by springs. This spring is a simulation of the phosphodiester bond between them. As we are currently using a spring, we can determine the opposite force and energy of the elongation by using the Hooke's Law, $E = ½ k \Delta X^2$, in which k represents the stiffness of each helical parameter (k_{shift}, k_{slide}, k_{role}, k_{till}…) and ΔX stands for the final elongation from the equilibrium state of the six helical parameters. Knowing that, the energy provided between two base pairs can be obtained following the previous reported formula. [19, 58, 59] The final energy of the whole system will be equal to the sum up of all energies, being represented as $E_{total} = \Sigma E$. [19]

The stiffness is a constant value which differs from each base pair of nucleotides and defines the amount of the opposite force the system has to exert against its equilibrium state. Stiffness can be obtained from different methods, but the bending stiffness obtained for our computational simulation comes from a previous Molecular Dynamics (MD) simulation by deriving the covariance matrix and later on inverting those values following Einstein's equation to recover stiffness matrices. The formula used is $\Theta = K_B T \subset^{-1}$ in which Θ corresponds to the stiffness, expressed in Newton/metres, K_B stands for the Boltzmann constant in m^2 kg s^2 K^{-1} and T is the absolute temperature in Kelvin (K). [19, 60, 61]

As previously described, high values of energy are associated with high values of stiffness and, as they're directly proportional, less spring elongation will be achieved. The opposite effect happens with low energy and stiffness values, in which high flexibility is obtained due to a high elongation rate of the spring. [19, 58]

3.4 Metropolis Monte Carlo algorithm

The Monte Carlo method is an algorithm used in computers that simply consists in integrating and operating over random numbers instead of a regular array. It takes into account the randomness and the statistics to generate and simulate systems or structures, so later on the total behaviour can be analysed. It's also applied in situations where the algorithm has got a condition and has to make a decision, with the output being yes or no. [62-65]

In our DNA simulation program, this algorithm is going to increase and decrease randomly a random number of our six helical parameter coordinates of a random base pair step. [116, 118] The energy of this newer configuration will be calculated and the code will check whether this new energy is smaller than the previous one without change. If the answer to the final value is yes, then the first value is changed for this newest one and the algorithm is repeated, but if the answer is no, then the same step is repeated with different random numbers. The Monte Carlo algorithm always pretends to obtain structures around the minimal energy configuration of the system, and for that reason the process is repeated as many times as wanted but decreasing the total energy value of the system in nearly every step, except for some conditions detailed in the simulation procedure. Finally, a lot of configurations of our structure are obtained which can later on be analysed through statistics. [66-68]

3.4.1 Statistics in the randomness: Obtaining the proper result

Statistics are someway really impressive, they let us obtain and understand the behaviour of a system from the randomness. One of the principles of analytics stands that if we have a lot of numbers and we do the mean of all of them, we obtain a result which is nearly exact, so that can be also applied to DNA simulations. If a lot of structures are generated due to the Monte Carlo equation, the mean of all the parameters from all the structures will provide us a final structure that resembles quite exact the properties of the examined DNA sequence. We conclude that the behaviour and the flexibility of a DNA helix can be determined from the randomness.

4. Simulation

The simulation part is involved in determining DNA flexibility by applying the Monte Carlo's algorithm in order to obtain an ensemble of representative structures, precisely with this algorithm we will obtain ten thousand representative structures of the examined DNA sequence. Later on, an analysis of the bending states will be done to determine whether we will have more chances to obtain high bending pieces or low ones. The sequence we examine will be 168 base-pairs long, 8x the single repeat of the test sequence. At this DNA length it is expected that the DNA, independent of the modification at the CpG step, is able to circularize, which is important for the experiment.

4.1 Procedure

For our simulation, DNA base-pairs are being represented as plain surfaces, and the phosphodiester bonds between two consecutive base pairs will be represented as springs. To run the Metropolis Monte Carlo algorithm, the initial energy of the whole system needs to be known. For that reason, as we are working with springs, we can obtain the energy between two base pairs by applying the Hooke's Law equation $E = ½ k \Delta X$. After previously knowing the stiffness constant from a Molecular Dynamics (MD) simulation between each consecutive base pairs and its current elongation, we can determine their applied energy against the equilibrium by executing the Hooke's law, but it has to be taken into account that, as DNA has helical coordinates, we will have six different equations with k_{shift}, k_{slide}, k_{role}, k_{till}… between two base pairs. The sum of all the energies from each single stiffness of the helical coordinates will provide us the total energy between two base pairs.

To determine the whole energy of the system we have to sum up the single energies of the 167 base pairs steps. We have 168 base pairs but as we are taken into account the energy of the string, there's one base pair at the end that won't have a spring, for that reason we only have 167 base pairs steps.

Once the total energy is calculated by following $E_{total} = \Sigma E$, we can apply the Monte Carlo's algorithm by writing small code in C, a programming language. Its principle can be divided in 4 steps which are:

1. First of all, helical coordinates are set for the initial position of each base pair steps. Total initial energy for each single base pair step is also set.
2. As previously described, the Monte Carlo's algorithm plays with the randomness, so now it will select a random base pair step from the 167 given, and change randomly a random number of the six coordinates we provided before. Just an example, it can select randomly the base-pair step 15, again randomly select three coordinates and change them by increasing or decreasing their value randomly.
3. From the obtained results, the total energy (E_{tot}) of the system is calculated again by applying the Hooke's Law.
4. As the energy of a system aims to reach minimum energy values as pointed by thermodynamics, we'll provide some conditions to our code so that our total energy will always be lower or equal to the previous one. In this condition, if the newer energy value is smaller than the E_{tot} before the change, the new configuration will be kept, otherwise, if it's bigger than the previous one, we'll keep the configuration before the change.

To avoid getting always to the minimum value of the system, we'll input a condition in which the Metropolis can take a random number between 0 and 1. If this number is bigger than $r > \mathrm{e}^{random_{(0,1)}} \frac{E_{tot\,new} - E_{tot\,old}}{E_B \cdot T}$, then the newer configuration whose energy is bigger than the one before change, will be accepted, so we'll have some picks when going around the minimum equilibrium value of the system.

This process is going to be done one hundred thousand times. In programming language, this process is called loop, just because the same procedure and steps are being done repeatedly. After the looping, we'll obtain a final configuration which we can transfer back from the six helical parameters coordinates to Cartesian coordinates in order to obtain a DNA structure which has been randomly created.

For our project, this algorithm has been done ten thousand times to obtain ten thousand different and independent DNA structures. After, the average of all the parameters from all the structures is done with the aim of obtaining a result which resembles quite exact the properties of the examined DNA sequence (Figure 13).

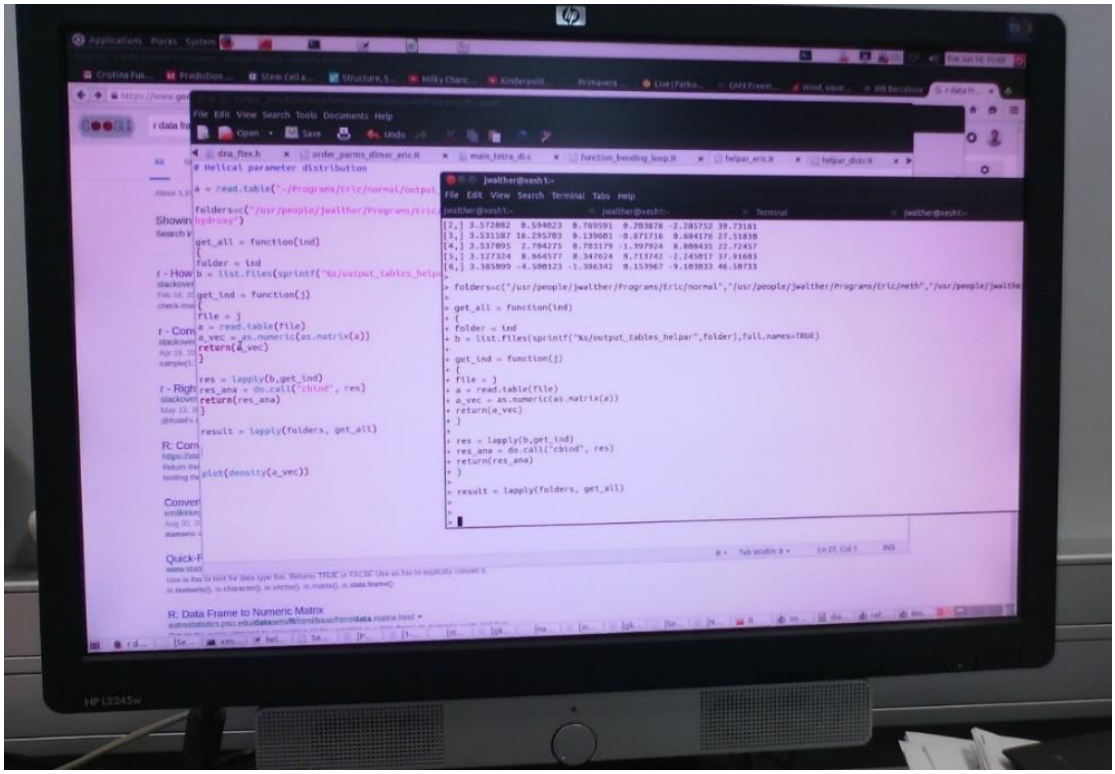

Figure 13. Process of running the code and the simulations. The Monte Carlo code was developed in C and run in the Linux terminal. Simulations were then run in a cluster, which has got a lot of processors so tasks are done within less time than using a computer. (Own authorship)

4.2 Results

After the Monte Carlo equation is done, we obtain ten thousand structures (Figure 14) which all came from the randomness of changing randomly the helical parameters of DNA and then transferring them back into the x, z, y axis.

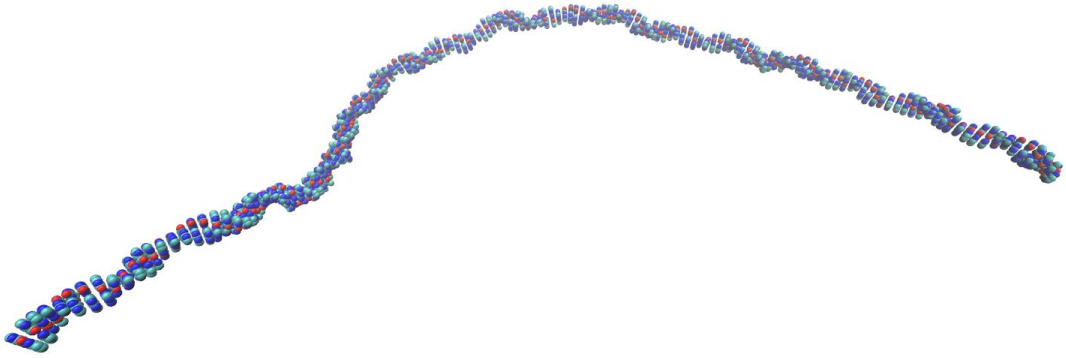

Figure 14. Simulation of naked-B DNA structure. This is one of the ten thousand individual representations obtained from the simulation, which represents a random structure due to using the Monte Carlo algorithm, which modifies randomly the helical parameters. (Own authorship)

4.2.1 Determination of stiffness and DNA flexibility

To determine our main topic of study, DNA flexibility, we can take a close look at the data provided for the different stiffness values for each model. By knowing that there is an inverse relation between stiffness-flexibility, the higher the stiffness is, the lower the flexibility will become. The first conclusion is that methylated and hydroxymethylated DNA seems to be less flexible than our reference DNA as their stiffness values are higher than the ones of the reference DNA. To confirm it, we have compared the stiffness values of the six helical coordinates (shift, slide, rise, tilt, roll, and twist in order of appearance in the tables).

For a correct analysis of the tables, the diagonal values from the left to the right need to be taken as the proper numbers to make the comparison.

Comparing both methylated and hydroxymethylated DNA, higher stiffness values are attributed to the hydroxymethylated due to the fact that near all the diagonal values are negative, indicating higher values over the methylated ones. Otherwise, there is an extreme positive value and a couple of them aren't very negative, so we can't get a clear conclusion between both DNA modifications, but regarding the contents of the table, we must suggest that hydroxymethylation is associated with a decrease of flexibility due to higher stiffness values over the methylated ones (Figure 15).

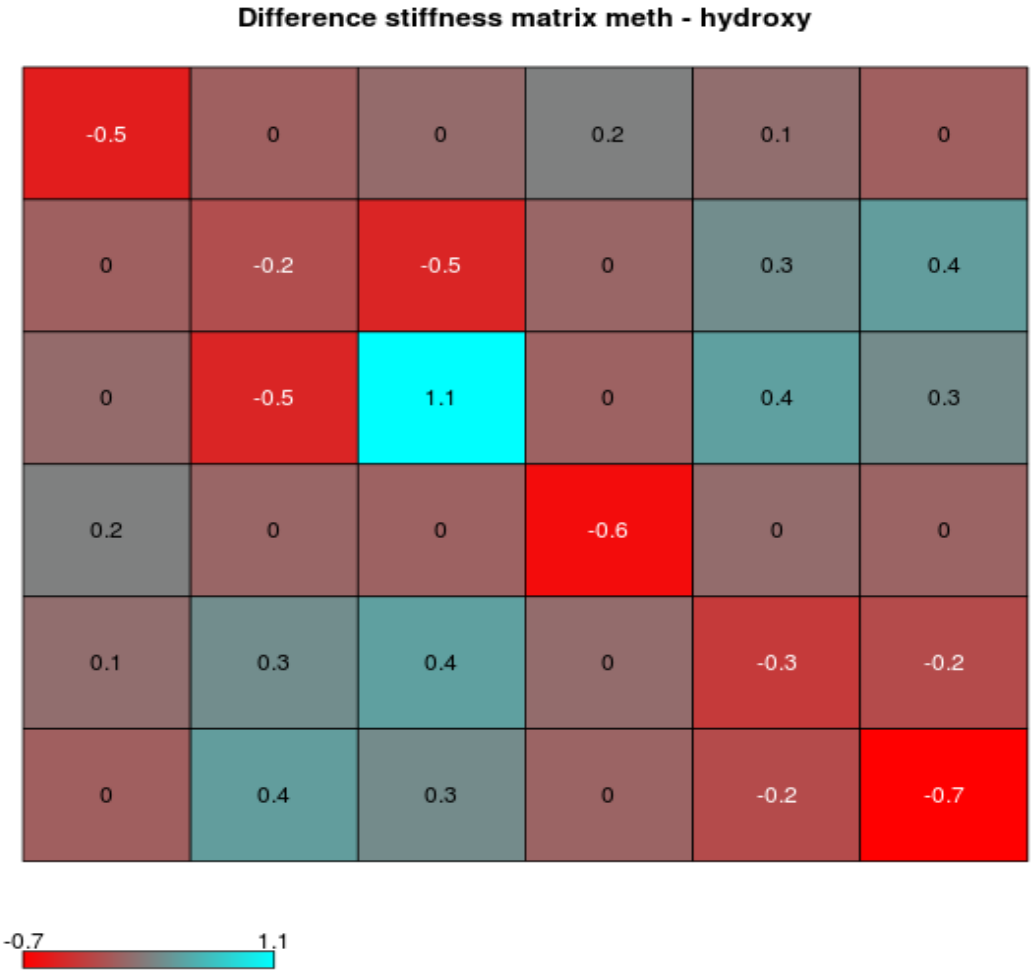

Figure 15. Difference stiffness between methylated DNA and hydroxymethylated. According to the legend, in red there are represented the most negative values (higher hydroxy values) and the most positive are represented in white (higher methylated values). It's found that hydroxymethylated has got higher stiffness values. Otherwise, there are also positive ones when looking at the table, so no conclusion can be obtained. (Own authorship)

Looking at the relation between reference and methylated DNA, all diagonal values seem to be higher in methylated DNA than in the reference because all diagonal values in the table are negative. There is a maximum negative value and the other ones are quite negative, too. Knowing that, we can determine that methylated DNA is less flexible than the reference one. Regarding our results, we must say that methylation on DNA is a modification that leads to a decrease of flexibility and, as a result, a reduction of probability to form a nucleosome core due to its low capability to wrap around histones (Figure 16).

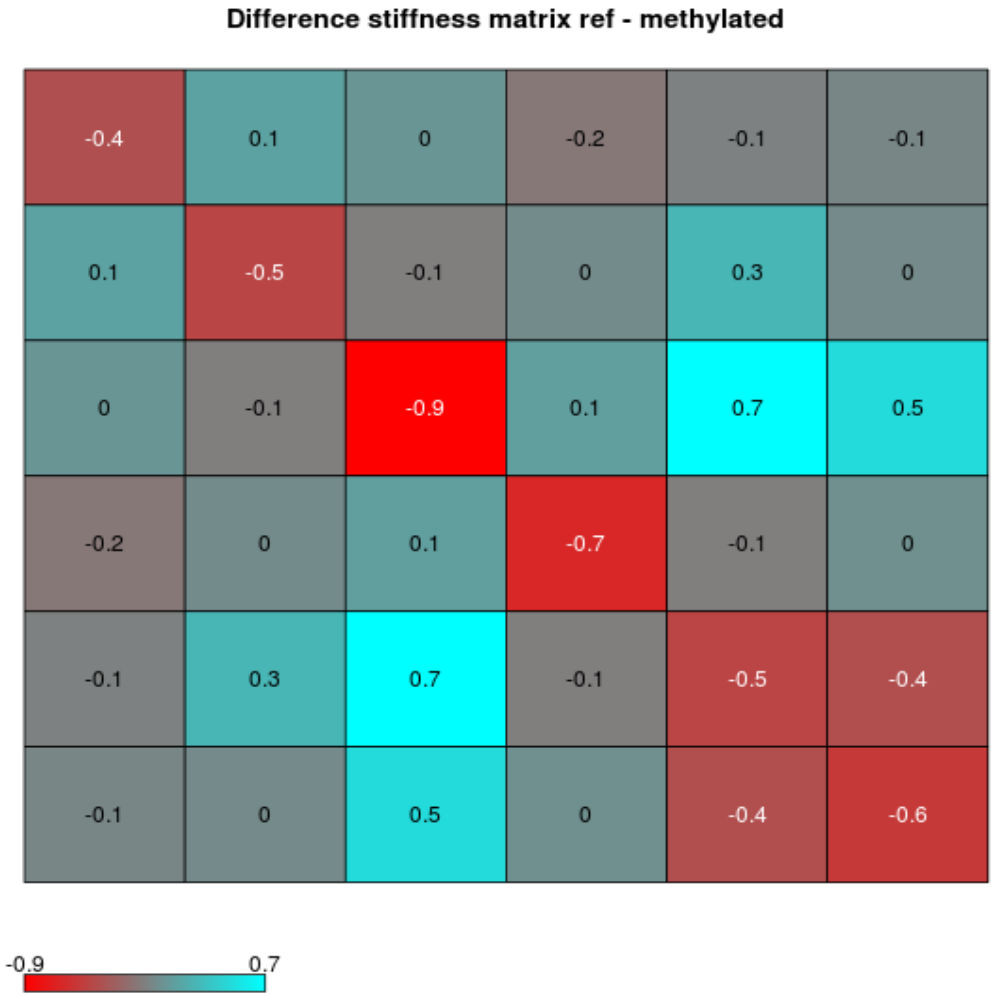

Figure 16. Difference stiffness between our reference and the methylated DNA. According to the legend, the most negative values are represented in red (higher methylated DNA) whereas the most positive ones are represented in white (higher reference). It's found that methylated DNA has got higher stiffness values than reference DNA due to the negative diagonal values, so it's less flexible than our DNA without modifications. (Own authorship)

Regarding at both reference and hydroxymethylated DNA, nearly all the diagonal values are negative, so hydroxymethylated DNA has got higher stiffness values than our reference DNA with no modifications. It is remarkable that there is a low positive value even though there are two maximum negative ones. As a conclusion, hydroxymethylated DNA corresponds to a lower DNA flexibility due to higher stiffness values than reference DNA, taking into account the inverse proportion between stiffness and flexibility (Figure 17).

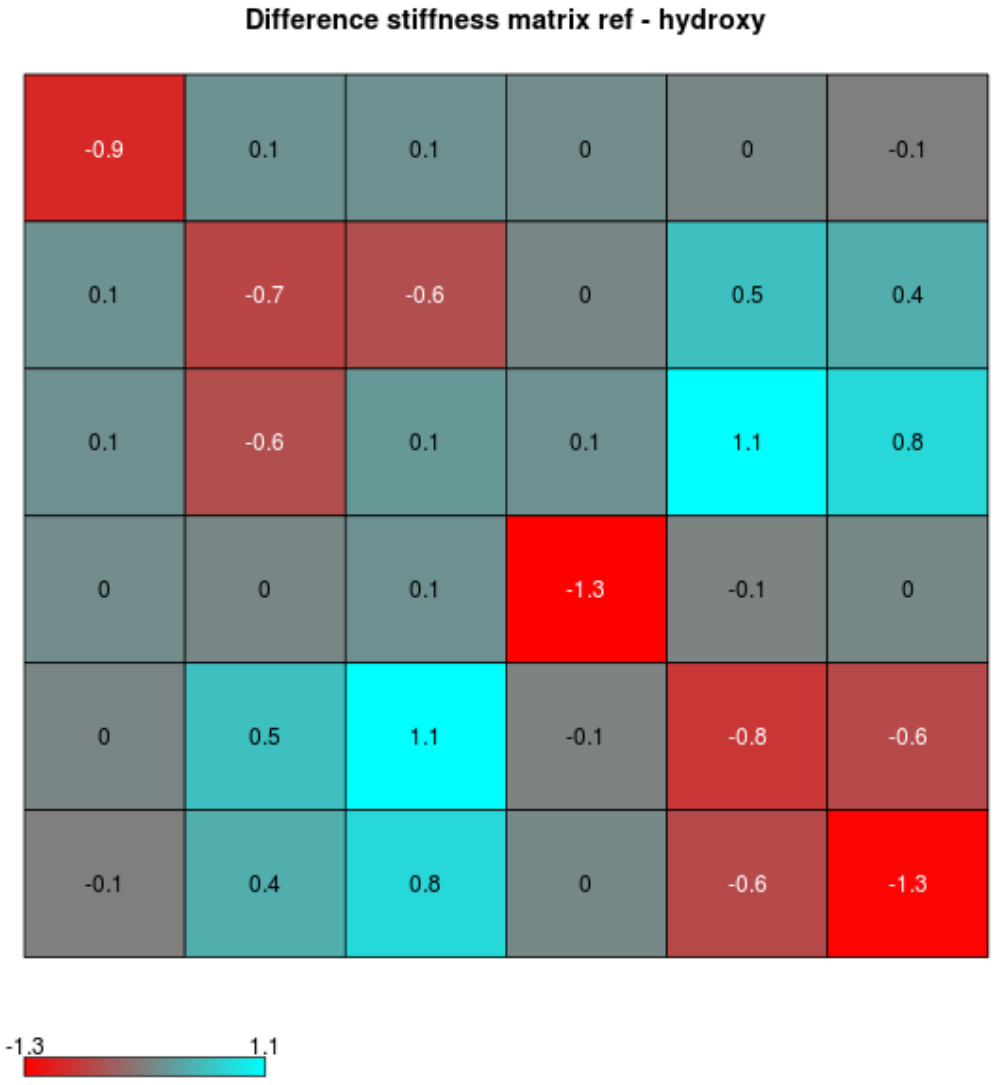

Figure 17. Difference stiffness between reference DNA and hydroxymethylated DNA. As the legend indicates, red colour is attributed to the lower values (higher hydroxymethylated) whereas blue colour is related with the most positive ones (higher reference). Taking a look at the table, hydroxymethylated DNA has got higher stiffness values than reference DNA, attributing less flexibility. (Own authorship)

After comparing all three DNA's, the conclusion obtained according to the flexibility is that Reference > Methylated ≈ Hydroxymethylated DNA. There is not a clear conclusion between methylated DNA and hydroxymethylated DNA as both of them have got positive and negative values and the rest values are very closed to each other.

We could not study the flexibility with the simulation of the formylmethylated DNA due to the absence of previous stiffness values, so no simulation could be done, but we do have results from the experimental part.

4.2.2 Study of the equilibrium local curvature

Another aspect that we can study with the data obtained from the simulation is the equilibrium parameters of the DNA from the CpG base-pair. Knowing the initial coordinates at equilibrium of the DNA will let us determine if the local curvature is higher or lower and later we can apply our results from the study of stiffness in order to describe the bending of the different DNA samples.

Regarding the following table, equilibrium values for the methylated and hydroxymethylated DNA are very close. Otherwise, we can describe that methylated DNA has got more shift at equilibrium, so it's going to produce a slightly higher local curvature due to its movement along the x axis respect to the other base pair. On the other hand, hydroxymethylated DNA has got higher twist values at equilibrium, so we will expect a slightly higher torsional rotation from the previous base pair at its equilibrium, which is not involved in changes in the local curvature. (Figure 18).

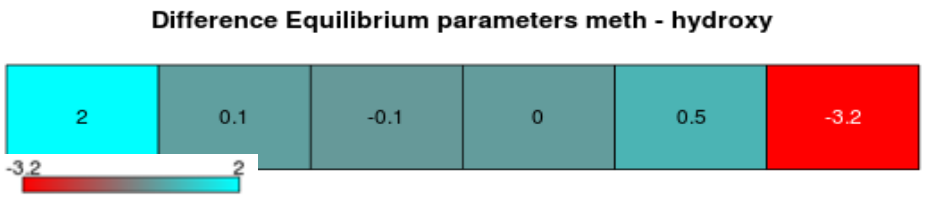

Figure 18. Difference in equilibrium parameters between methylated and hydroxymethylated DNA. In red it's described the most negative values (higher hydroxyl) and in blue the positives one (higher methyl). It's found that hydroxymethylated DNA will have more twist whereas methylated DNA will have more shift. (Own authorship)

Doing the same comparison between reference and hydroxymethylated DNA we can see that also equilibrium values are very close altogether but there are slight differences. Hydroxymethylated DNA has got a very high value of roll, which means that it will start with a predetermined bending respect to the next base pair, so its local curvature will be high. About the reference DNA, we can see high twist values, which are involved in high movement along the y axis. Otherwise, base pairs won't move a lot from each other, resulting in a low local curvature (Figure 19).

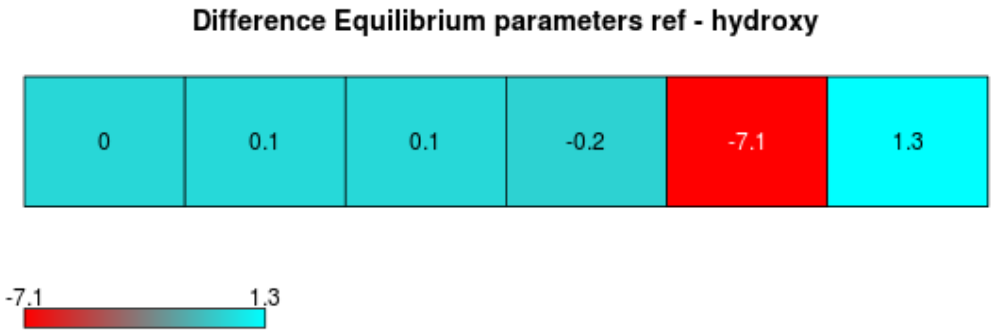

Figure 19. Difference in equilibrium parameters between reference and hydroxymethylated DNA. Red colour described the maximum negative values (high hydroxymethylated) whereas blue colour indicates positive values (high reference). The most remarkable threat is the high roll of the hydroxymethylated, which will provide high local curvature, and the high twist of the reference. (Own authorship)

Finally, we can compare both reference and methylated DNA. All equilibrium coordinates are very close but there are more differences than in the previous comparisons. Here, we can observe higher methylated values in both shift and roll, which are both related with the production of a high local curvature in equilibrium because the base pair moves through the x-axis and then it rotates (bending) respect to the following base pair. If we look at the methylated DNA, it also has got high twist values like in the previous comparison, so we can determine that it will rotate a lot in the y-axis but it's going to remain with a low curvature value, as it doesn't move through the x axis like the hydroxyl or the methylated ones (Figure 20).

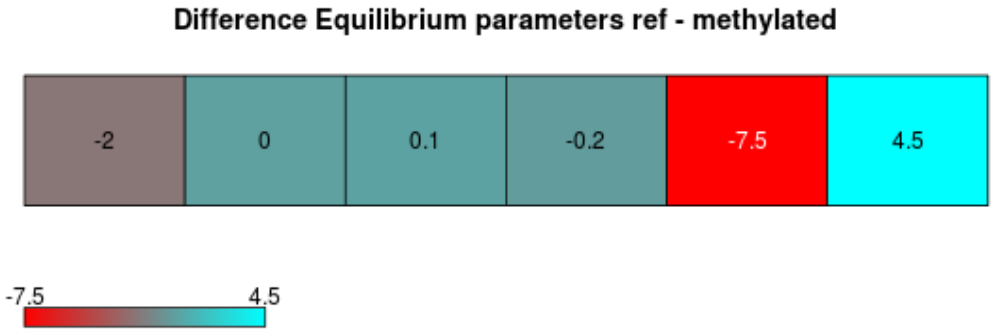

Figure 20. Difference in equilibrium parameters between the reference and the methylated DNA. Following the legend, in red there are represented high minimum values (high methylated) and positive values in blue (high reference). It's found that methylated DNA has higher roll and shift values, so a higher local curvature than reference DNA, which has a high twist value, related with high torsion on the y-axis, but no local curvature. (Own authorship)

As a conclusion of the previous comparison, methylated DNA and hydroxymethylated DNA will have a high local curvature due to the high values of the roll and shift respect to reference DNA, which are related with a far x-axis movement and rotation respect from the following base-pair. Regarding the reference DNA, it's shown that is has a high twist, which is related with a high torsional rotation but it's not going to have a high local curvature as it's not moving far from the x, y axis, so a low local curvature will be attributed to the non-modified DNA.

4.2.3 Analysis of the bending

After describing the stiffness and the local curvature, we can provide an analysis of the probability of getting pieces with high or low bending by looking at the CpG site, where we applied the modifications.

From the relation between the local curvature and the stiffness, a graphic has been done for comparison. Regarding the graphic, we can describe that methylated and hydroxymethylated DNA have higher local curvature values than the reference one, namely from around 15 to 25 degrees, which is something that we could expect from previous analysis (read *Study of the equilibrium local curvature*).

As we know from previous analysis, high local curvature attributed to methylated DNA and hydroxylmethylated DNA and a high level of stiffness will result in pieces with high bending due to the fact that our representative spring will be compressed (no movement) by high stiffness values and the coordinates make that both consecutive base-pairs are far from each in the x-axis, so high bending pieces will be obtained as shown in the figure (Figure 21).

According to reference DNA, from previous analysis we determined that it has low local curvature and low stiffness values, which will result in getting more pieces with low bending but with a slight separation between two consecutive base-pair. This conclusion is achieved due to that fact that the spring won't be so much compressed, so it will be more elongated and the low local curvature results in the base-pair not moving a lot from the x-axis, so lower bending pieces will be obtained (Figure 21, 22, Table1).

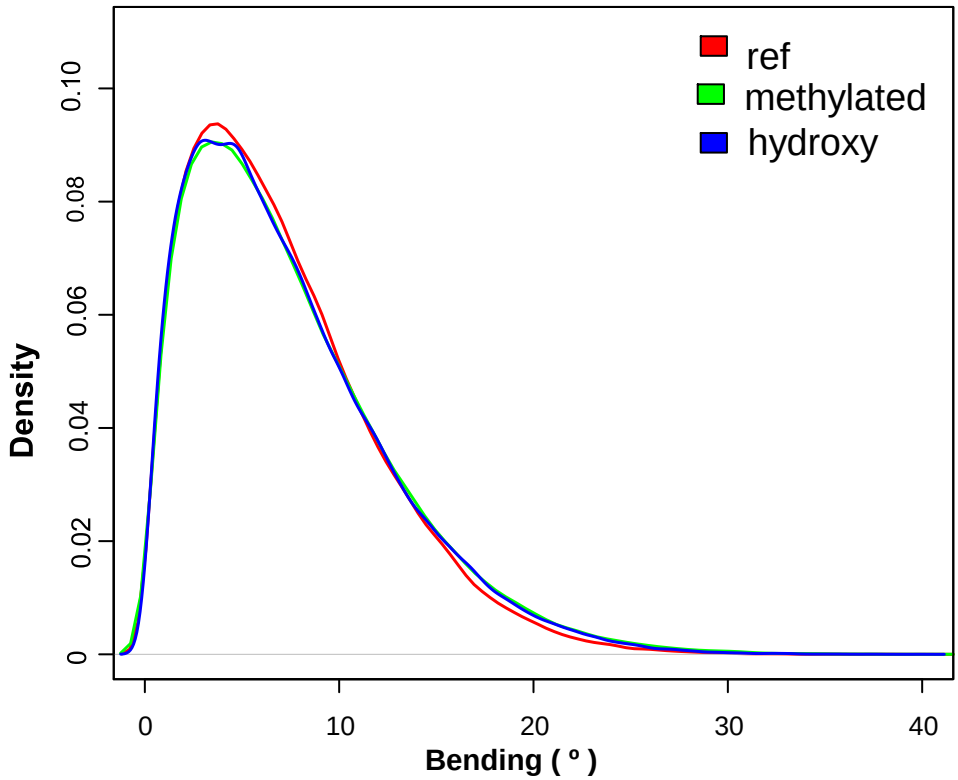

Figure 21. Relation between bending and density. As described in the legend, red colour corresponds to reference DNA, green for methylated DNA and blue for hydroxymethylated DNA. It appears that methylated and hydroxymethylated DNA have got more pieces with higher local curvature at around 15-25 degrees. (Own authorship)

Table 1. Relation between local curvature-stiffness and the chance of getting high or low bending pieces in the different DNA simulation tests done. High local curvature and high stiffness are involved in obtaining more pieces with high bending due to a high compression of the spring and a slight movement in the x axis of one base pair respect to the following one. This affirmation fits with our methylated and hydroxylmethylated description. Low stiffness and low local equilibrium values are related with low bending as bases aren't compressed, so they can move a lot but around low values from the equilibrium, so more pieces with low bending pieces are obtained. (Own authorship)

	LOCAL CURVATURE IN EQUILIBRIUM VALUES		
STIFFNESS		↗	↘
↗	Less bp-pieces with low bending More bp-pieces with high bending	More bp-pieces with low bending Less bp-pieces with high bending	
↘	Less bp-pieces with low bending More bp-pieces with high bending Distribution of bending broader than above	More bp-pieces with low bending Less bp-pieces with high bending Distribution of bending broader than above	

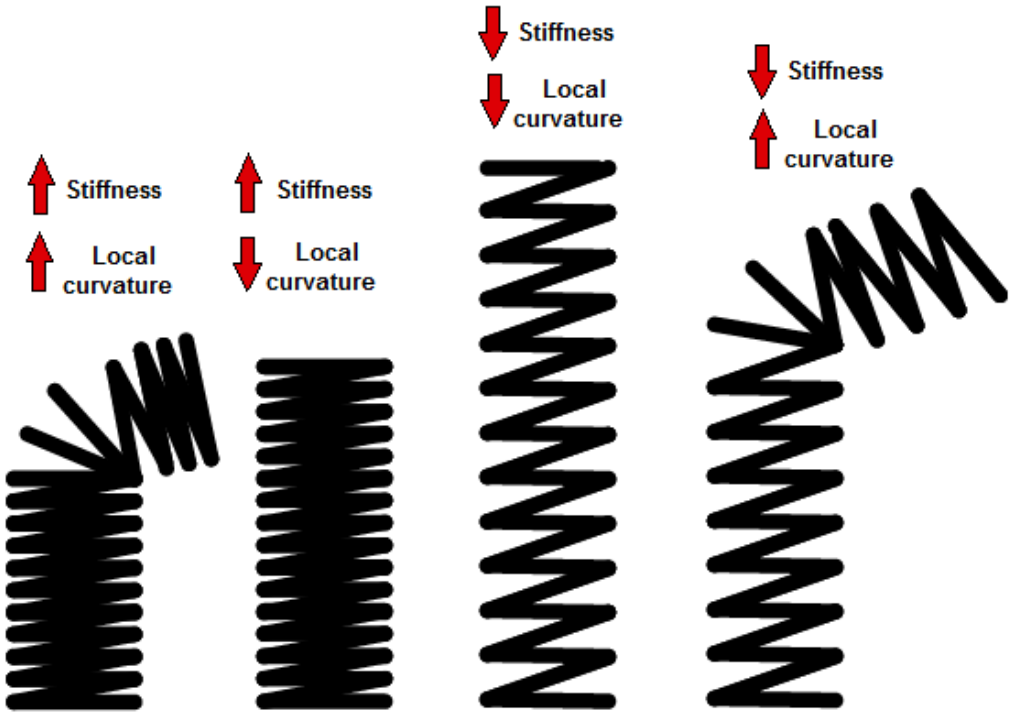

Figure 22. Representation of the probability of getting high or low bending pieces by analysing the relation between stiffness and local curvature in equilibrium. As seen in the picture, high stiffness and high local curvature leads to more high bending pieces, whereas high stiffness and low local curvature increases the formation of low bending pieces, as a decrease of the base-pairs movement respect to the following one. Low levels of stiffness and low local curvature are related with low bending pieces as bases can move a lot but they can't because of the need of being closed to their equilibrium value. Low stiffness and high local curvature is associated with getting more high bending pieces due to the high capability of movement along the x-axis. (Own authorship)

5. Experiment

The experimental part involves the making of a circularization assay in order to quantify the circularized DNA in our different conditions (reference, methylated, hydroxymethylated and formylmethylated). For doing that, we need to follow the circularization protocol and later on do a 1st dimension and a 2nd dimension gel in order to obtain a visual representation of our circular and linear DNA. Then we will perform a quantification using the Image J software and compare the results between our different samples in order to determine their flexibility and connect the results with a biological role.

5.1 Procedure

Table with the relation between Procedure-steps and Theory location in the manuscript.

	Brief description	Theory location
1. Phosphorylation	Add –P groups to 5' ends of DNA fragments	**2.4 Theory behind circularization assays [Page 13]**
2. Annealing	Cool down DNA to form double strand structures	
3. Ligation	DNA ligase used to connect DNA fragments through phosphodiester bonds.	
4. Ethanol precipitation	Na+ attaches to DNA and leads to precipitation. Purification step	
5. 1st Electrophoresis gel	Separate DNA fragments depending on their size	**2.5.1 1st Dimension Electrophoresis [Page 14]**
6. 2nd Electrophoresis gel	Separate circular-linear DNA	**2.5.2 2nd Dimension Electrophoresis [Page-15]**

5.1.1 Phosphorylation

We have the following 8 oligos (100 µl of stock in each one):

- Reference DNA Forward strand (Ref For)
- Reference DNA Reverse strand (Ref Rev)
- Methylated DNA Forward strand (Met For)
- Methylated DNA Reverse strand (Met Rev)
- Hydroxylmethylated DNA Forward strand (hMe For)
- Hydroxylmethylated DNA Reverse strand (hMe Rev)

- Formilmethylated DNA Forward strand (fMe For)
- Formilmehtylated DNA Reverse strand (fMe Rev)

TOTAL: 50 µl (x 8 tubes)

Figure 23. Different aliquots with 100 µl of each one. We have two samples of each DNA, as one of them contains the forward strand and the other one the reverse one. (Own authorship)

1. Take 1 nmol of oligo (10 µl of 100 µM stock solution).
2. Add 5 µl of 10X Buffer. The quantity of Buffer added comes to the calculation that if our final solution will contain 50 µl and as it is 10X diluted, we will need 5 µl for our solution. It provides the ATP and the right conditions for the enzyme to be functional.
3. Add 31 µl of water (H_2O) to the volume.
4. Add 4 µl the T4 Polynucleotide Kinase (T4 PNK) at 10 u/µl (New England Biolabs). The T4 PNK phosphorylates the 5'-end of the DNA molecule.

Figure 24. Different single strand DNA samples. Each Eppendorf contains one of our DNA samples (reference, methylated, hydroxyl-methylated DNA, formyl-methylated) as shown in the picture. This DNA samples have been treated with buffer, water and a kinase which will phosphorylate the 5'end of the strands. (Own authorship)

Figure 25. Materials used for the phosphorylation process. At the top left, an ice bucket to conserve our DNA samples. At the bottom left, micropipette tips with different sizes depending on the quantity that we have to take. At the right, micropipettes, measurement instruments used for taking small quantities (from 0,1 µl to 200 µl) of our samples. Each micropipette has got a corresponding tip to work with. (Own authorship)

The phosphorylation process, which involves the addition of a phosphate group at the 5'-end of the DNA molecule thanks to the kinase is done so that later on phosphodiester bonds with following DNA molecules can be done thanks to the free –OH group located at the 3'end.

5.1.2 Annealing

1. Mix the Forward and the Reverse strands of each pair of DNA samples.
2. Put the mixed samples 2 min. at 95ºC.
3. Leave it to cool down slowly until temperature reaches more or less room temperature.

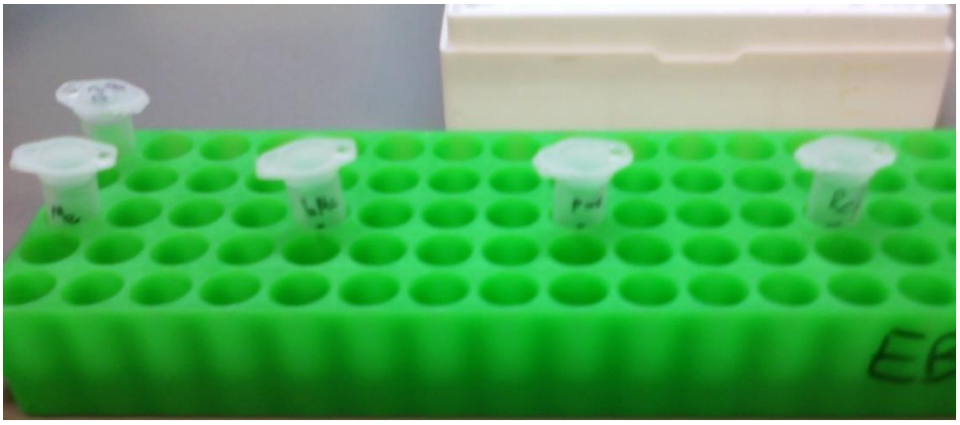

Figure 26. DNA samples with both forward and reverse strands mixed. (Own authorship)

Figure 27. DNA samples being hot at 95ºC to denaturalize DNA. Hot temperature is added to ensure that DNA is single stranded (denaturalized) and later on it is cooled down to obtain the double strand structure due to the formation of the H-bonds as a result of the temperature decrease. (Own authorship)

What is happening through this process is that we heat up our samples to ensure that we have single stranded DNA and avoid the formation of hydrogen bonds. Later on, the decrease of temperature is going to induce the formation of hydrogen bonds between the nucleotides (2 between A-T and 3 between C-G) until all bonds will be formed and as a result we will have 100 µl of double stranded DNA (for each DNA).

5.1.3 Ligation

1. 0,2 nmol of each sample is required from the initial 1 nmol volume. As the equivalence between µl and nmols is known (100 µl is equal to 1 nmol), 20 µl from each DNA sample will be taken from the annealing reaction.
2. Add 2,5 µl of 10X T4 DNA ligase Buffer. 2,5 µl are determine as that the overall solution will be of 25 µl and we have a 10X Buffer, diluted 10 times, so 2,5 µl are needed. This buffer provides the optimal conditions for the DNA ligase.
3. Add 1,5 µl of water (H_2O).
4. Add 1 µl of T4 DNA ligase at 400 u/µl (New England Biolabs). A ligase is an enzyme that is essential in the circularization assay as it form the phosphodiester bond between the extremity 5'-phosphate of one DNA fragment and the 3'-OH of another one.
5. Incubate 1h at 22ºC.

Ligation is the most important step because here is produced the circularization of the DNA, the more flexible DNA is, the less fragments it will require to form circles. Not all the molecules will ligate effectively, so the product of the reaction will contain circularized DNA and linear DNA. Not all the circularized molecules will be completely closed (both DNA strand ligated) and we will obtain also some circularized molecules in an open conformation, in which, one of the two strands won't be ligated.

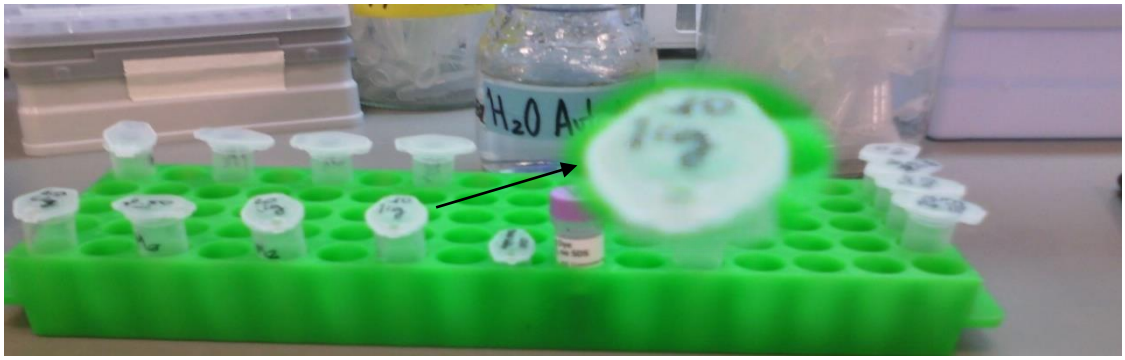

Figure 28. Ligation samples at the left and DNA samples at the right. The Eppendorf with the ligation reaction is shown. (Own authorship)

5.1.4 Ethanol precipitation

1. Add three times the volume of the samples (3·V) of ethanol 100% and a tenth of the volume of the sample of Sodium Acetate (NaAc 3M pH 5,2) (V/10). In this case, we add 2,5 µl of NaAc 3M pH 5,2 and 75 µl of Ethanol 100% to 25 µl of Ligation reaction.
2. Keep overnight at -20ºC.
3. Centrifuge the samples 15 min at 4ºC at 13000 rpm.
4. Remove the supernant and wash the pellet with 50 µl of Ethanol 70%. This wash is done in order to remove the excess of salt.
5. Centrifuge 5 minutes at 4ºC at 13000 rpm.
6. Air dry the pellets to evaporate the excess of ethanol in the Eppendorf tubes.
7. Resuspend in a solution of 10 µl of 1X loading buffer (NEB), which is a mix of 2,5% Ficoll®-400, used for creating brighter and tighter bands when compared to glycerol loading dyes, 10 mM EDTA, 3,3 mM Tris-HCl, 0,02% Dye 1, which serves as tracking, 0,001% Dye 2 and pH 8,0 at 25°C.

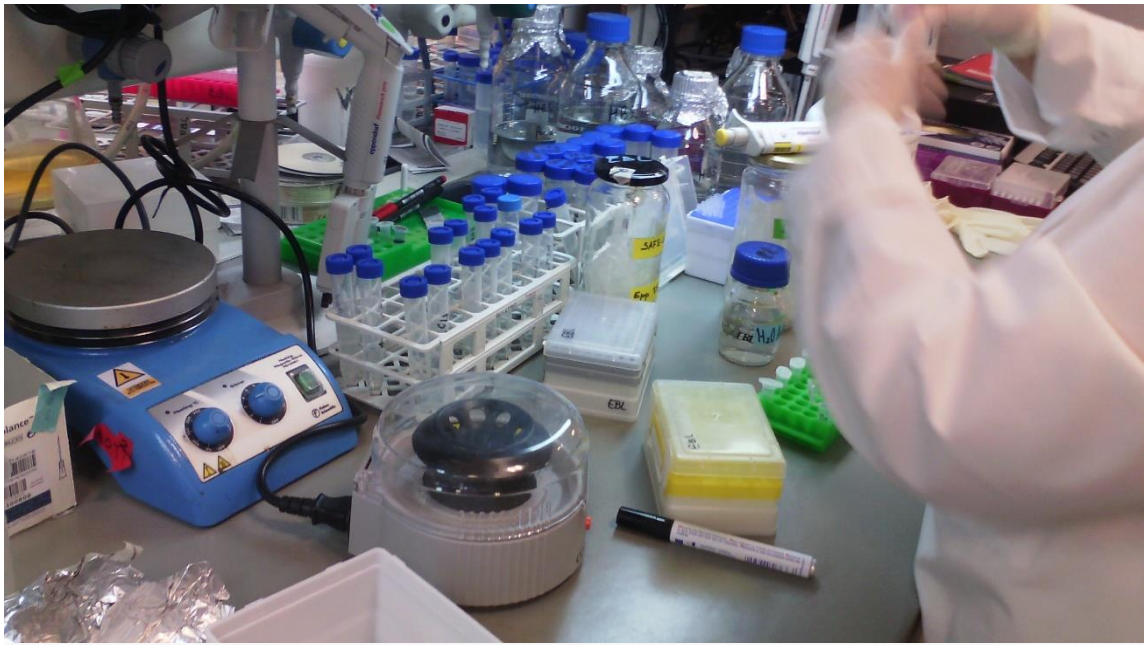

Figure 29. Process of doing the ethanol precipitation. At this moment, the drying process is being done. It is a very accurate process because the pellet looks white and it remains at the bottom of the Eppendorf. A spin is given to the samples with the centrifuge in order to obtain the pellet at the bottom of the tubes and know where they are even if they are not clearly visible. (Own authorship)

5.1.5 1st dimension electrophoresis gel

1. Prepare the solution:

- TBE 10X 1 ml
- Acryl/Bis 40% 1,25 ml
- H2O 7,6 ml
- APS 10% 50 µl
- TEMED 10 µl

TOTAL: 10 ml

APS is used as a catalyser to increase the speed of the polymerization whereas TEMED is the component that allows the polymerisation.

2. Take the solution from the tube and slowly pour it between the glass plates. We use plates of 1 mm distance between them.

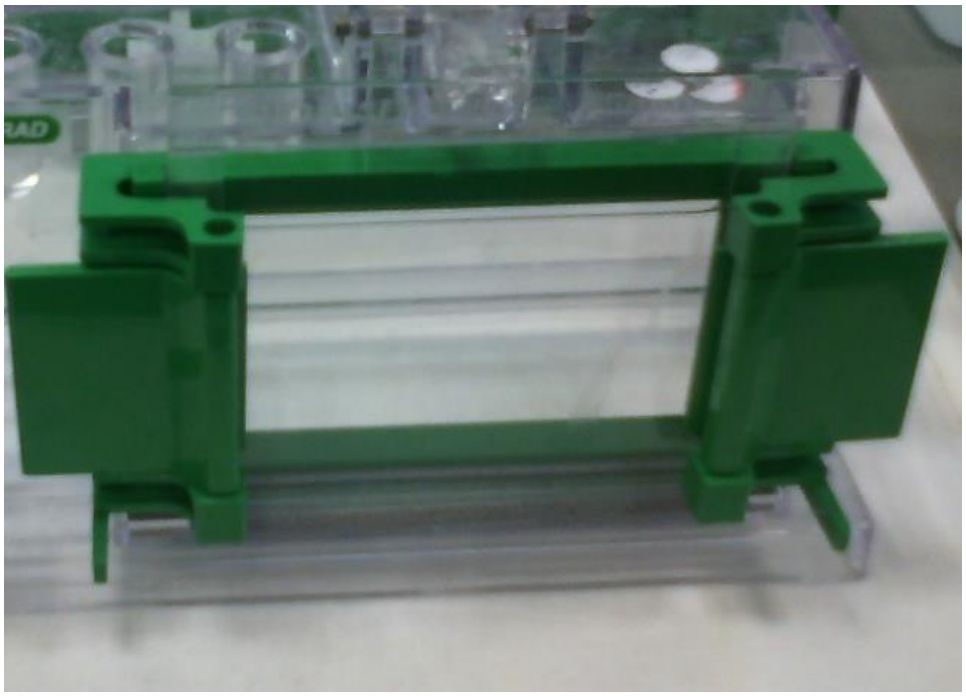

Figure 30. Electrophoresis gel system. The two plates are located each together and attached. Our solution is poured in the 1mm space between the plates. Gel will be polymerised inside. (Own authorship)

3. Wait 30 minutes until the gel is fully polymerised.

4. Put the gels into the tank of the electrophoresis apparatus and fill the tank with running buffer (TBE) in order for the electricity to pass through the system allowing DNA to migrate to the cathode (positive pole).
5. Load the four DNA samples into the wells of the gel. Leave an empty well between samples and load the 1 kb molecular weight marker (Invitrogen) in the first well. Into the empty wells, put some loading buffer to avoid the samples to diffuse in the gel.
6. Run the electrophoresis at 30 V for 3 h 40 min.

Figure 31. Migration of the DNA samples. After the samples are loaded into the wells, voltage is applied so our negative samples can migrate towards the positive pol. (Own authorship)

7. Stain the gel for 15 min in Ethidium Bromide. We dilute a stock of 10 mg/ml with TBE 1x to a final concentration of 50 µg/ml, which is a dilution of around $20 \cdot 10^3$ times.
8. Remove the Ethidium Bromide and put the gels in water to eliminate the excess.
9. Leave the gel in water for 30 min.

This is the normal protocol used for doing the 1st dimension electrophoresis gels. As a result, our DNA has migrated and bands are obtained which represent the different size of the molecules. If they are really big, they will be stuck at the top because the holes of the gel won't let them pass, whereas if molecules are small, they will migrate more easily and they will be found at the end of the gel. When migration has finished, the bands are not visible. For that reason Ethidium Bromide, which is a colorant that intercalates into DNA, is applied so we stain our

DNA and it can be observed and analysed with UV rays into a dark chamber. We prepared two gels, as we had 7 samples, 2 of reference DNA, 2 of methylated DNA, 2 of hydroxymethylated DNA and 1 of formylmethylated DNA, so one of each was charged in each gel, except for the formylmethylated DNA one, that as it was difficult to observe from previous experiments, we decided to do only one.

Figure 32. Process of placing our DNA gels to the dark chamber. The dark chamber provides darkness, so UV lights don't have any type of interference and our samples can be monitored. (Own authorship)

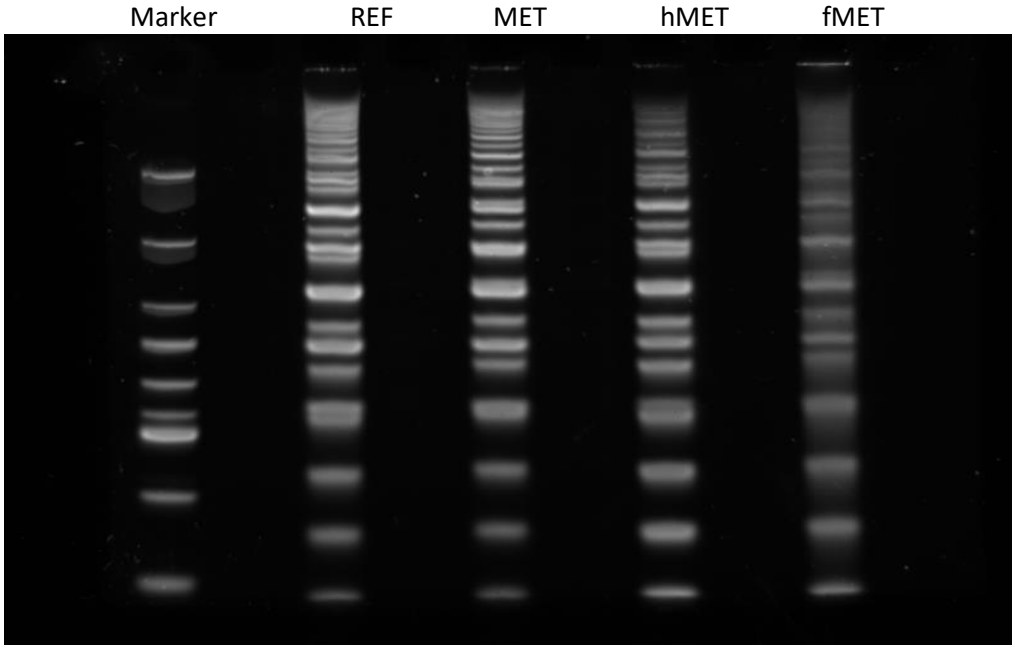

Figure 33. Visualization of our gel bands with our stained DNA samples through UV light. From left to right: the molecular weight marker (Marker), reference (REF), methylated (MET), hydroxymethylated (hMET) and formylmethylated (fMET) DNA. (Own authorship)

5.1.6 2nd dimension electrophoresis gel

For the 2nd gel electrophoresis, 7 gels have to be prepared as we are loading one lanes from the 1st dimension gel per gel; so as we loaded 7 DNA samples, the same amount of gels need to be prepared now.

1. Prepare the gel:

- TBE 10X 1ml
- H2O 6,93 ml
- 40% Acryl/Bis 2 ml
- APS 10% 50 µl
- Chloroquine (50 mg/ml) 50 µl
- TEMED 50 µl

TOTAL: 10 ml

Chloroquine attaches to DNA fragments and allows the resolution of circularized DNA by unwinding a bit the double helix and decreasing the coiling of the topoisomers, so they can be resolved better.

2. Take the solution from the tube and slowly pour it between the plates. We use plates of 1,5 mm distance between them so there is space to introduce the 1 mm thick lane from the 1st dimension gel.

   Add isopropanol at the surface of the gels, because they need to be complete flat so the bands can stick and successfully migrate.
3. Wait 30 minutes until the gel is fully polymerised
4. Cut each band from the first gel and place it horizontally on the top of the 2nd gel. A bit of loading buffer can be added to help the band slide between the 2 plates.

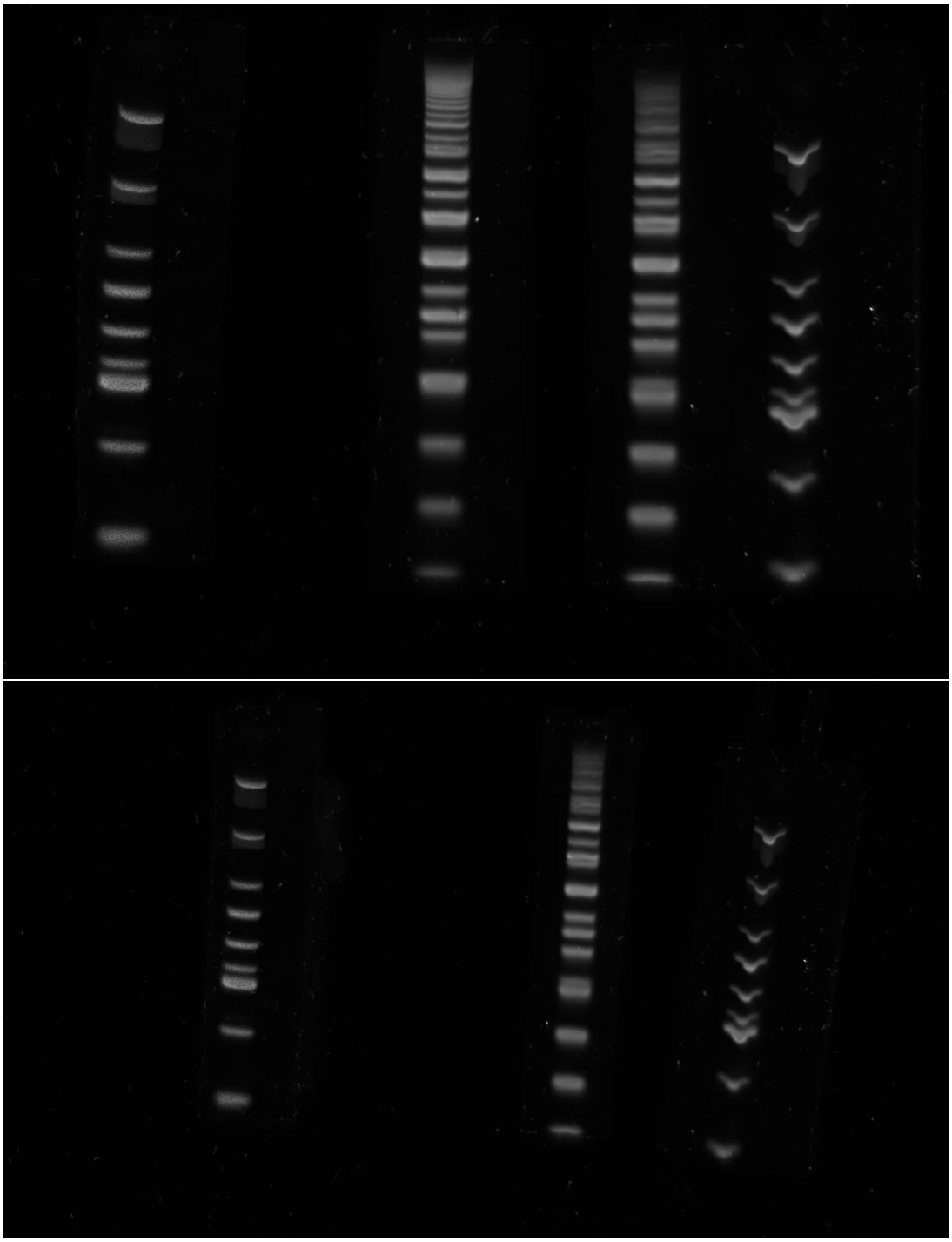

Figure 34. Visualization of cut the bands of the gels. To make the 2nd dimension gel, it's needed to load the different bands of our DNA samples into another electrophoresis system, so the bands of the gel have to be cut in order to do that. As four samples have been loaded in the 1st dimension gels, we will obtain four bands to run. (Own authorship)

5. Put the seven gels into the dispositive and fill it up with running buffer.

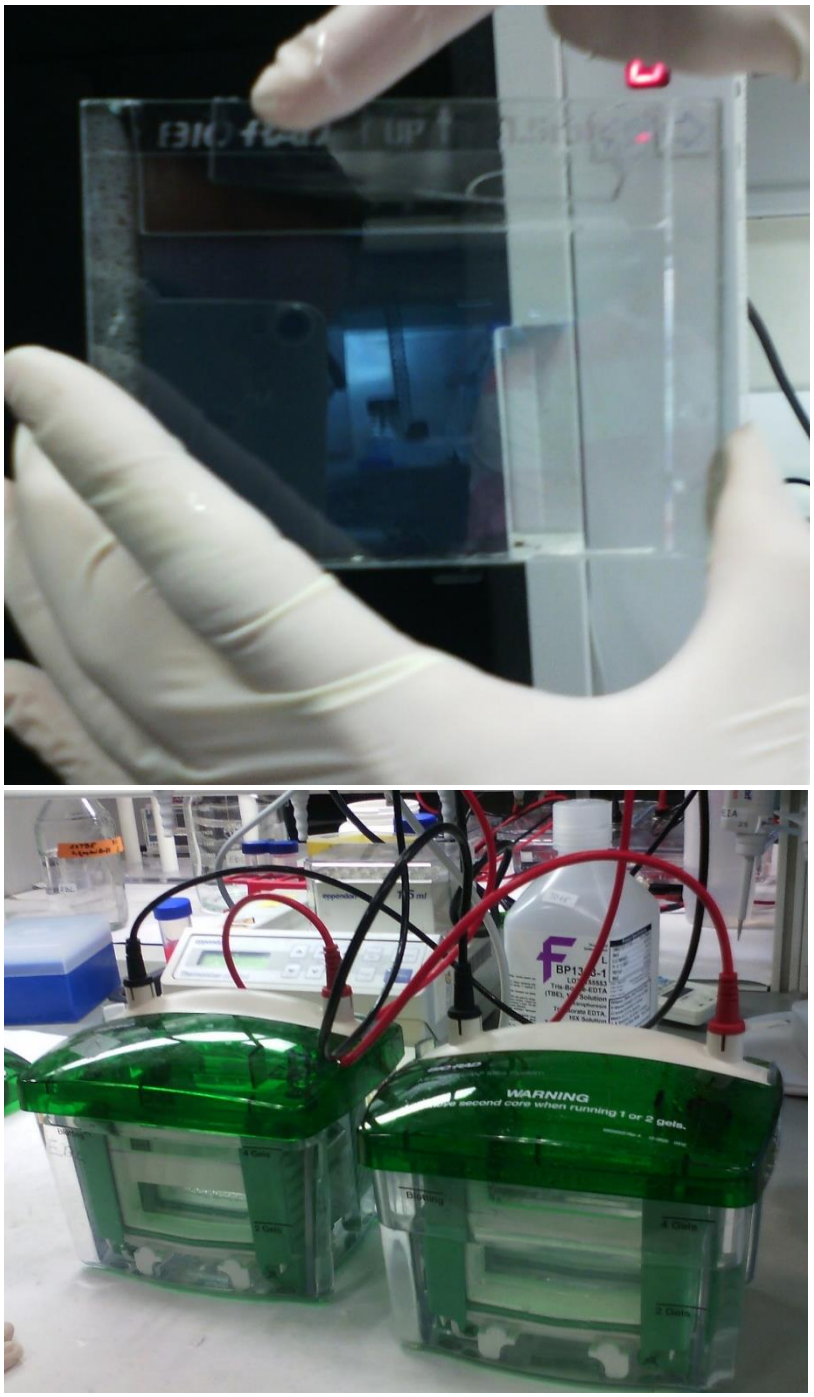

Figure 35. Loading the bands to the 2nd dimension gel and placing them into the electrophoresis apparatus. The four bands have to be put horizontally in order to be in contact with the gel and avoid interferences. This four gels are then placed in the electrophoresis apparatus. (Own authorship)

6. Add purple loading buffer at the top to check that migration is happening.
7. Migrate overnight (16 hours) at 18 V.
8. Stain the gel for 15 min with Ethidium Bromide. We dilute a stock of 10 mg/ml with TBE 1x to a final concentration of 50 µg/ml, which is a dilution of around $20 \cdot 10^3$ times.

9. Remove the Ethidium Bromide and put the gels in water to eliminate the excess.
10. Leave the gel in water for 30 min.

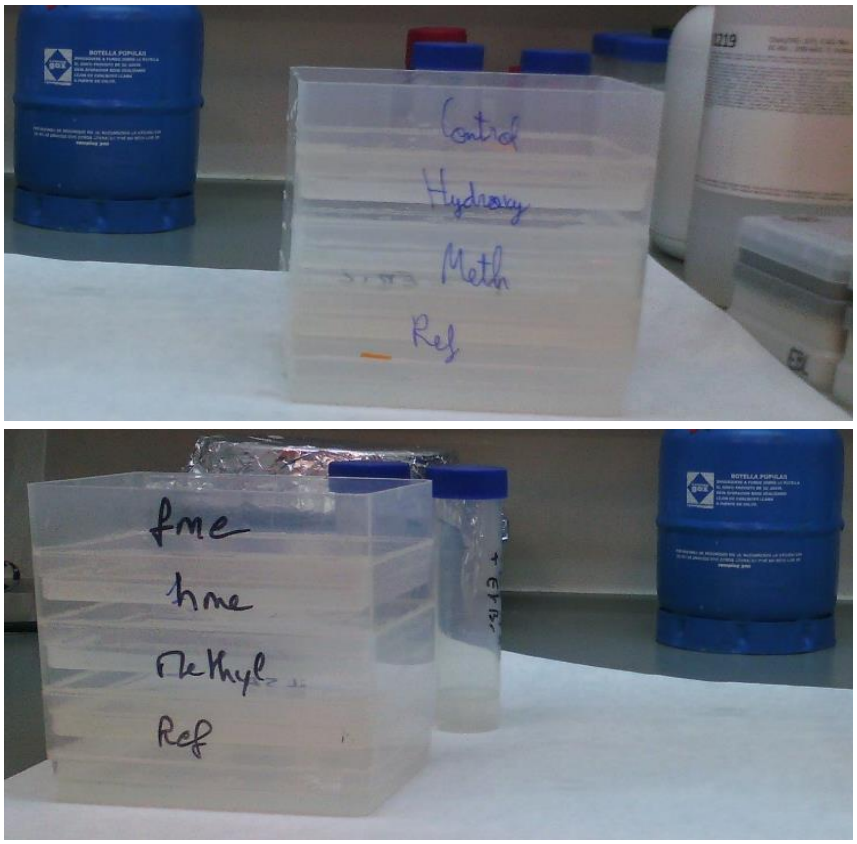

Figure 36. Gels left in water to remove the exceed of Ethidium Bromide. Ethidium bromide is a very strong stainer, so we need to dilute its concentration after staining the gels, for that reason some washes are done. (Own authorship)

The 2nd dimension protocol has been followed for doing the second electrophoresis gel. Here, molecules are separated depending on their molecular weight and on their conformation so circularized DNA can be separated from linear DNA, so as a result we obtain a type of curves with white dots on it and a black background after checking the gels in the dark chamber with UV rays.

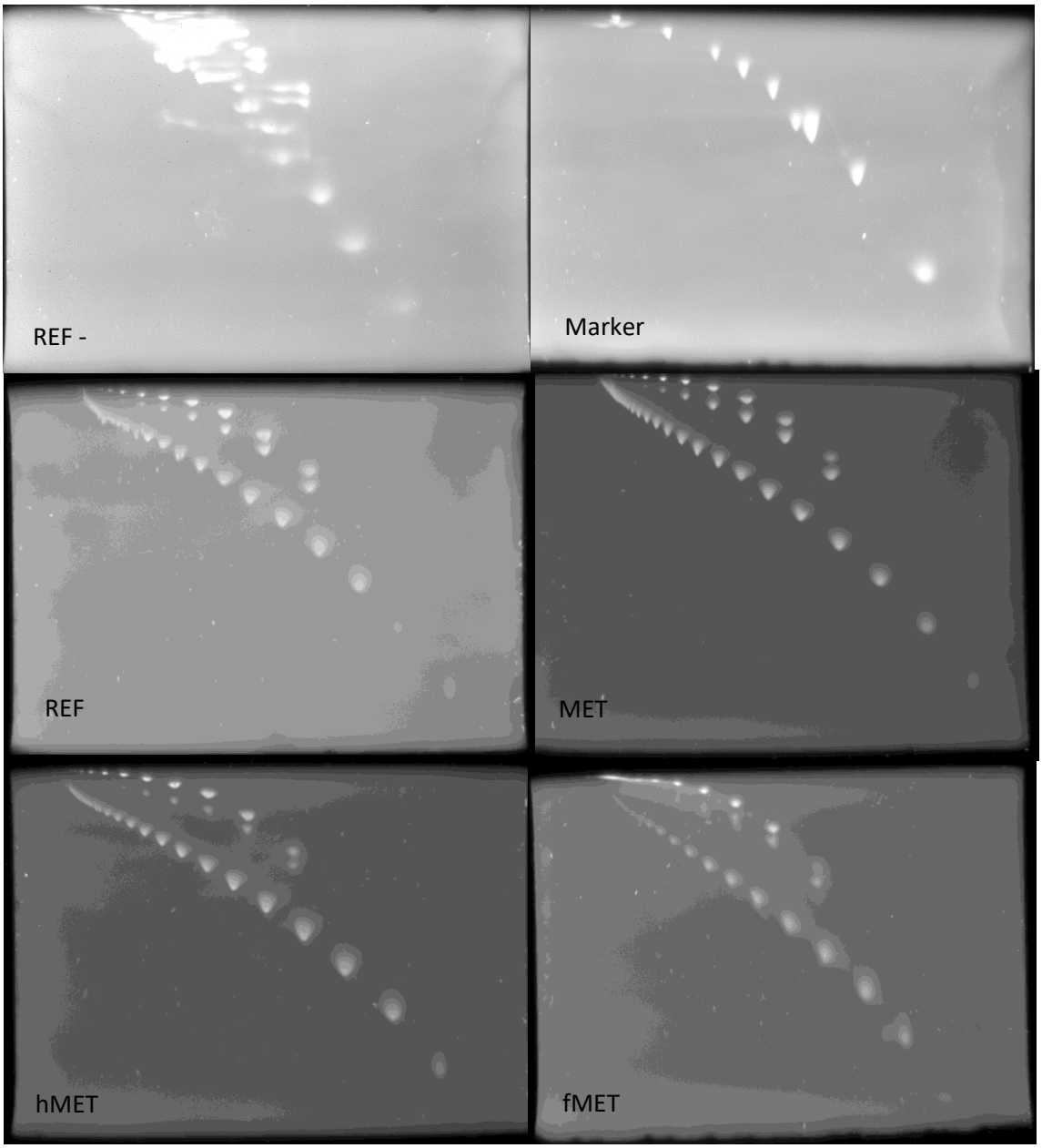

Figure 37. Visualization of the 2nd dimensions gels in the dark chamber with UV light. It is represented a bad gel migration, the marker, reference, methylated, hydroxymethylated and formylmethylated DNA in order. Circularized DNA is found at the top of the gel whereas linear DNA represents the lower line. The quantifiable dots start at 105 bp. It is known by comparing the migration dots with the marker DNA and the individual bands obtained from the 1st gels. It is remarkable that bad migrations can happen as seen in the first picture, so not all experiments work at their first attempt, and that the loaded marker has only one line, which is found for the linear DNA. (Own authorship)

5.2 Results

After the 2nd dimension electrophoresis gel, a quantification is performed to determine the index of DNA molecules that have circularized during the process over the linear ones, which did not circularize. This information is needed to describe the flexibility of our DNA samples.

5.2.1 Quantification of DNA flexibility

For the quantification, a program called Image J has been used, which allows us to quantify the spots of the gel pictures by looking at their integrated density. We knew from the 1st dimension gel that our 2nd dimension gels ranged from 105 bp an increasing each spot. As we didn't see any quantifiable 105 bp spot of circularized DNA in our samples, we decided to start quantifying from 126 and take 5 measurements, until 210. After obtaining the measurements of the integrated density, we have done a closed versus linear comparison in order to determine the quantity of circularized molecules and the tendency of our different samples to make circles (Figure 38).

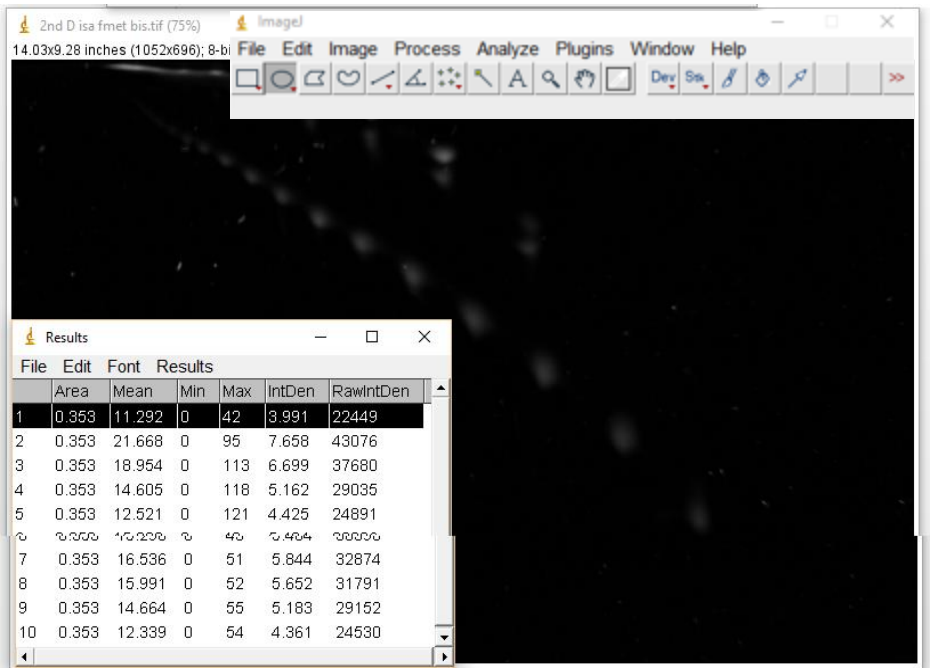

Figure 38. Image J quantification process. After obtaining the image of the gel, it has to be processed through Image J in order to obtain white dots in a dark background. Small circles are done with the tool that cover all dots and later one it's set to measure to obtain the Integrate Density of each dot, the value that determines the quantity of circles and linear DNA at a certain molecular weight. (Own authorship)

After obtaining the data, we see that our reference DNA is more flexible due to the fact that it has more circularized DNA values than the opened ones. According to the graphic, we can describe that methylated DNA is less flexible than reference due to lower values in circular DNA, hydroxymethylated DNA is even less flexible and we expect a similar behaviour for formylmethylated DNA, but we can't get a precise conclusion just to the lack of measurement techniques and a low resolution of the gel (Figure 39).

C/L	Ref	Met	hmet	fmet
126	1,65	1,21	0,69	1,35
147	1,80	1,46	1,11	0,76
168	1,66	1,36	0,94	0,84
189	1,30	1,09	0,85	1,00
210	0,97	0,85	0,68	0,99

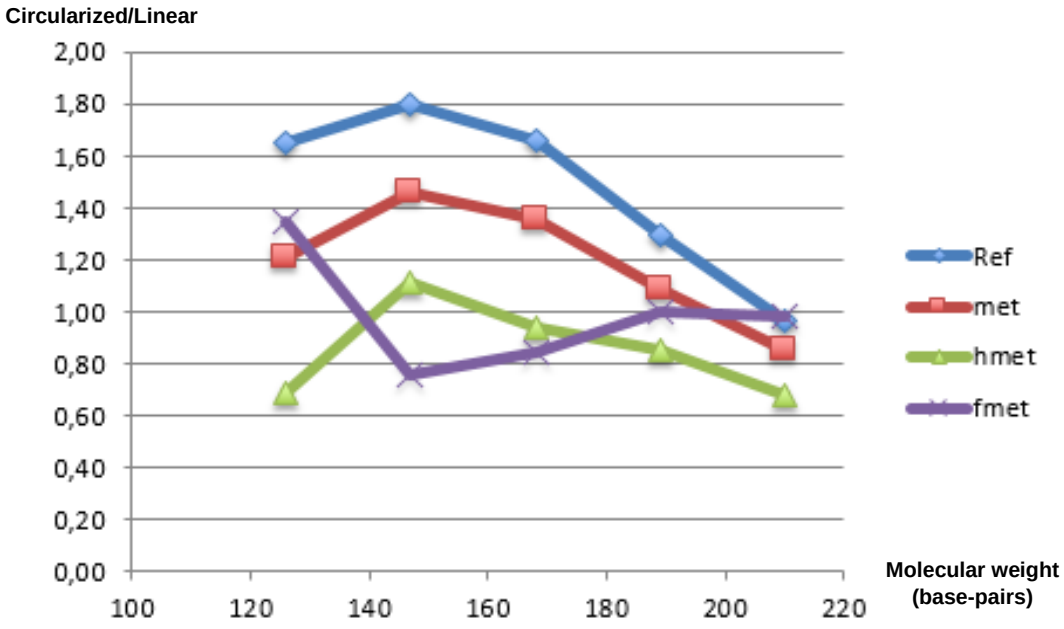

Figure 39. Graphical representation of the flexibility analysis. According to the legend, blue colour is attributed to reference DNA, the red one to methylated, green for hydroxymethylated and purple for formylmethylated. X-axis represents the length of DNA fragments; y-axis shows the ratio of circularized versus linear DNA. Regarding the graphic, reference has got higher values due to higher circular DNA over the linear, so it's more flexible because it tends to circularize more. It decreases for the methylated one, even more for the hydroxymethylated and it's not clear the result obtained from the formylmethylated, so further experiments need to be done. (Own authorship)

As a final result of the experiment according to the flexibility we concluded that Reference > Methylated > Hydroxymethylated ~ Formylmethylated. We did not study precise values, just the overall flexibility of our DNA molecule in the different modifications we applied. More data is required to analyse the formylmethylated DNA, but we would suggest that different variables (time, temperature) affects the behaviour of formylmethylated DNA and its final resolution, for that reason we cannot see a proper tendency even when it's compared with previous circularization assays.

5.2.2 Analysis of closed/opened circles

After doing the experiment, we saw that previous results suggested that there could be a relation between flexibility and the capability of DNA to form more closed circles or more opened ones. This happens because not all DNA close both strands when circularized, probably one of the strands gets closed but the other one remains opened, but it still circularizes.

To determine the relation between flexibility and the chance of obtaining more opened or closed circles, the same procedure as the one previously described for the circularization quantification has been done with the different spots of the circularized DNA gel using Image J. Analysing the graphic we can see that reference DNA has got higher closed circles than hydroxymethylated and the formylmethylated due to slightly higher values on closed over opened. It seems that methylated DNA forms more close circles even than reference.

Regarding the results obtained from the study of the flexibility, we can determine that there is a correlation between flexibility and quantity of closed circles respect to the opened ones. Reference DNA, which is more flexible, tends to make more closed circles when circularizing whereas hydroxymethylated DNA and formylmethylated DNA tend to form more opened circles. There is no explanation on why we get more closed circles with methylated DNA, as it should have more opened ones because it is less flexible than reference, so we suggest that it is related with the chemical properties of the methyl group, which change when being in contact with determined agents present in the experiment, increasing the chance of closing when the molecule is able to circularize, but more studies need to be done in order to solve it (Figure 40).

o/c	Ref	Met	hmet	fmet
126	0,85	0,46	1,61	0,60
147	1,21	0,76	2,27	1,93
168	1,58	1,05	3,18	3,00
189	1,57	1,26	3,29	
210	1,92	1,64		

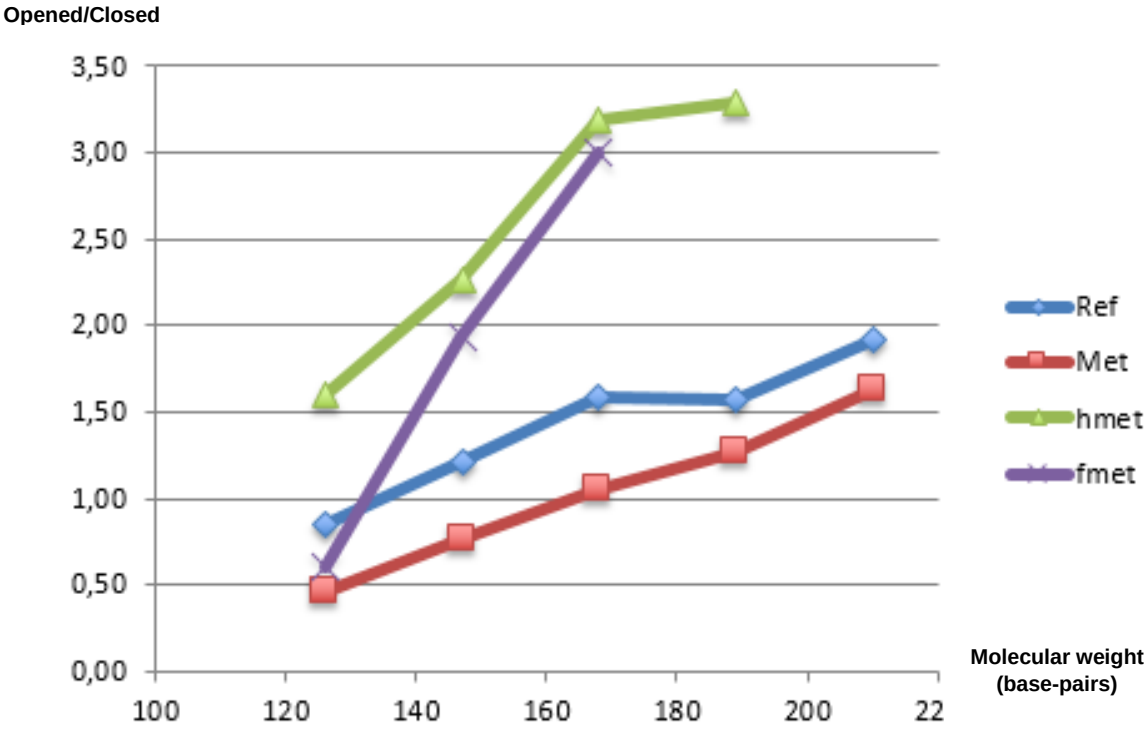

Figure 40. Graphical representation of opened circles over closed ones. Regarding the legend, reference is represented in blue, methylated DNA in red, hydroxymethylated in green and formylmethylated in purple. X-axis represents the length of the DNA fragment; y-axis shows the ratio of opened versus closed circular DNA. It's found low values at the reference, which means high values of closed circles over the overall opened ones. The opposite happens with hydroxymethylated and formylmethylated. This patterns fit with the current knowledge that reference DNA is more flexible, so it forms more closed circles than the others. There's no current explanation on why methylated DNA closes more than reference DNA while being less flexible than it.

6. Comparison between experimental and simulation studies

Once getting the results of the experiment and the simulation, we can compare both of them in order to obtain a conclusion. The first thing that we have to take into account is that in our simulation we have looked at the physical properties of the nucleotides and the bending force produced by the stiffness and the local curvature, which are physical assignments, whereas in the experiment we have looked at the dynamics and kinetics of DNA.

In both simulation and experimental studies we have reach the same conclusion, which is that methylation is a modification that decreases DNA's flexibility, and the pathway of demethylation seems to even slightly decrease DNA's flexibility. With the simulation we lead to the hypothesis that hydroxymethylated DNA decreased even more DNA's flexibility than the methylated one as it had slightly higher stiffness values, but we couldn't get with a conclusion until we checked the experimental results, which clearly confirmed the presence of a hydroxyl group decreases DNA's flexibility even more than the methyl group.

Some differences that we obtain from both studies are that through the simulation we see that the flexibility difference between reference and methylated DNA is bigger than the one between methylated and hydroxy-methylated. Furthermore, when observing the experimental results, we observe the same pattern of difference between them. There is the same flexibility difference between reference and methylated than between methylated and hydroxymethylated, which is something that is not compatible when regarding the simulation results. Probably it happened because in the simulation we are taking a close look at the physical properties and in the experiment we are only taking into account the capability of circularizing without looking at possible effects of the chemical DNA modifications.

7. Development of novel epigenetic drugs: Applications of the results obtained in this study

Epigenetic drugs have recently been in clinical trials to treat some of the most devastating diseases, gaining special interest in the cancer field by changing histone post-translational modifications and therefore inhibit oncogenes or express genes (by acetylating histones) that downregulate cell division, leading to amazing results. Otherwise, this drugs are a bit imprecise because nucleosome formation is increased or decreased but it's not known which genes are going to be affected, being probable that some key genes for the cell are disturbed and leading to cell death. For that reason, novel epigenetic drugs which are more efficient and more precise need to be developed.

Why is this project relevant in this field? We now know that cytosines can be methylated and that this modification decreases DNA flexibility and therefore increasing gene expression. We suggest taking advantage of our results as an approach to develop novel epigenetic drugs involving a system made of CRISPR with a methyltransferase. As an example, if we have the gene *p53* (it acts as a checkpoint to treat DNA damage and leads to apoptosis) disrupted by some mutation, we will be more likely to suffer from DNA alterations and therefore have cancer. If we are able to target another gene involved in those processes which is not being expressed and express it, the cell will undergo in apoptosis and the death of the cancerous cell. There, our plan is to develop a CRISPR system recognizing this gene involved in cell division and attach a methyltransferase to the system which will methylated the cytosines located in that gene, decreasing at that point DNA flexibility and nucleosome formation, consequently increasing the expression of that gene, which will theoretically make the cell to undergo apoptosis. So, which advantages may this novel system have in front of the current available epigenetic drugs? Precision, as only the fragment of interest is being treated, and accessibility, because cytosines are more accessible to modifications than the amino acid residues of histones.

We have to point out that this is only a suggestion of possible applications of this study, but further experiments and trials need to be done in order to rely in this method.

8. Discussion

We have shown that DNA methylation could play an important role in gene expression by affecting DNA flexibility. As seen in both simulation and experimental assays, we can conclude in terms of flexibility that Reference > Methylated DNA > Hydroxyl-methylated DNA. We don't have simulation data for formyl-methylated DNA. We can hypothesise that these modifications may have a biological importance because they contribute to either increase or decrease nucleosome formation by changing DNA flexibility.

As we did not have the stiffness values of formylmethylated DNA from previous simulations, we could not compare anything with the experimental results. Moreover, bad resolution of the formylmethylated DNA didn't allow us obtain a proper and valid conclusion, so we suggest working on that field in future projects as well as redefining the protocols to avoid chemical incompatibilities with the chemical groups added in the DNA (methyl, hydroxyl, formyl).

In the experiment, we saw that there was the same difference pattern in flexibility between all DNA samples, whereas the simulation suggested that there was a close flexibility relation between hydroxymethylated DNA and methylated DNA than in the reference, so more precise quantification techniques or different protocols for getting a better resolution of the gels are required in order to explain this difference.

Furthermore, when studying the relation of opened versus closed circles of our DNA samples, we observed that methylated DNA used to do more closed circles than reference DNA, but this does not fit with our finding that methylated DNA is less flexible than reference, because it would have to form more open circles as a result of the lower flexibility. We suggested that it could be by a chemical disruption of the methyl group when being in contact with some solution within the process, but we should conduct further studies because previous results suggest the same pattern.

Some further projects and open questions that actually remain unknown could be done in order to determine the biological processes involving the requirements of the cell to do the passive demethylation pathway, because it would be more logic that an enzyme would directly remove the methyl group from the DNA. Moreover,

as studies have shown that some of the demethylation modifications, such as the formylmethylated DNA, are found in embryonic stem cells or promoters, so we suggest that evolution has favoured this system through generations, although the reason still remains unknown. Another possible object of study would be to understand whether DNA methylation can also be applied to nucleosomal DNA (when the nucleosome core is already formed) or if linkers are the only ones being affected.

9. Acknowledgements

I would like to acknowledge the Crazy About Biomedicine program for providing me an amazing opportunity to develop this research work, to all the amazing people in the IRB Barcelona who has supported me in Modesto Orozco's lab and in Structural and Computational Biology's Lab.

Specially, I would like to thank two amazing people, Isabelle Brun Heath, the Lab Director, for giving me the chance to perform an experiment and instruct me during the time I was at IRB Barcelona. The other person is Jürgen Walther, who is my IRB tutor and has bet on me all the time that I could afford the project and helped me a lot with both simulation, the project and making my staying the most comfortable possible.

Finally, I would like to thank my tutor, who has helped me a lot in the process of structuring the work and finding the right way on how to say it.

As a conclusion, I acknowledge everybody who has made possible this project. Without any of you the final result would not have been possible.

10. Bibliography

1. Sarai A, Mazur J, Nussinov R, Jernigan RL. Origin of DNA helical structure and its sequence dependence. *Biochemistry*. 1988;27(22):8498-8502.

2. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. The Structure and Function of DNA. 2002.

3.Dahm R. Friedrich Miescher and the discovery of DNA. *Dev Biol*. 2005;278(2):274-288.

4. Chargaff E, Lipshitz R., Green C. Composition of the desoxypentose nucleic acids of four genera of sea-urchin. *J Biol Chem*. 1952;195(1):155-160.

5. Dickerson RE. Definitions and nomenclature of nucleic acid structure components. *Nucleic Acids Res*. 1989;17(5):1797-1803.

6. Crick, F. H. C. & Watson, J. D. (1954). The Complementary Structure of Deoxyribonucleic Acid. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *223*(1152).

7. Mitchell D, Bridge R. A test of Chargaff's second rule. *Biochem Biophys Res Commun*. 2006;340(1):90-94.

8. Wang JC. Helical repeat of DNA in solution. *Proc Natl Acad Sci U S A*. 1979;76(1):200-203.

9. Drew HR, Wing RM, Takano T, et al. Structure of a B-DNA dodecamer: conformation and dynamics*. Proc Natl Acad Sci U S A*. 1981;78(4):2179-2183.

10. Reiter GF, Senesi R, Mayers J. Changes in the Zero-Point Energy of the Protons as the Source of the Binding Energy of Water to A -Phase DNA*. Phys Rev Lett*. 2010;105(14):148101.

11. Campos EI, Reinberg D. Histones: annotating chromatin. *Annu Rev Genet*. 2009;43:559-599.

12. Cooper GM. Chromosomes and Chromatin. 2000. Book. Consulted: 10/11/2016

13. Suganuma T, Workman JL. Signals and combinatorial functions of histone modifications. *Annu Rev Biochem*. 2011;80:473-499.

14. Kouzarides T, Aggarwal BD, Calvi BR, et al. Chromatin modifications and their function. *Cell*. 2007;128(4):693-705.

15. Strahl BD, Allis CD. The language of covalent histone modifications. *Nature*. 2000;403:41–5.

16. Kim H-J, Seol J-H, Han J-W, Youn H-D, Cho E-J. Histone chaperones regulate histone exchange during transcription. *EMBO J*. 2007;26(21):4467-4474.

17. Jin B, Li Y, Robertson KD. DNA methylation: superior or subordinate in the epigenetic hierarchy? *Genes Cancer*. 2011;2(6):607-617.

18. Razin A, Riggs AD. DNA methylation and gene function. *Science*. 1980;210(4470):604-610.

19. Pérez A, Castellazzi CL, Battistini F, et al. Impact of methylation on the physical properties of DNA. *Biophys J*. 2012;102(9):2140-2148.

20. Razin A. CpG methylation, chromatin structure and gene silencing-a three-way connection. *EMBO J*. 1998;17(17):4905-4908.

21. Lister R, Pelizzola M, Dowen RH, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*. 2009;462(7271):315-322.

22. Cheung P, Lau P. Epigenetic regulation by histone methylation and histone variants. *Mol Endocrinol*. 2005;19(3):563-573.

23. Allfrey V, Faulkner R, Mirsky A. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc Natl Acad Sci U S A*. 1964;51(5):786-794.

24. Martin C, Zhang Y. The diverse functions of histone lysine methylation. *Nat Rev Mol Cell Biol*. 2005;6(11):838-849.

25. Kohli RM, Zhang Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature*. 2013;502(7472):472-479.

26. Mosammaparast N, Shi Y. Reversal of histone methylation: biochemical and molecular mechanisms of histone demethylases. *Annu Rev Biochem*. 2010;79:155-179.

27. Nucleic Acids Book (http://www.atdbio.com/content/17/solid-phase-oligonucleotide-synthesis). Consulted: 26/06/16

28. Sekine M, Sekine M. DNA Synthesis Without Base Protection. In: *Current Protocols in Nucleic Acid Chemistry*. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2004:3.10.1-3.10.15.

29. McBride, L. J., Kierzek, R., Beaucage, S. L., & Caruthers, M. H. (1986). Nucleotide chemistry. 16. Amidine protecting groups for oligonucleotide synthesis. *Journal of the American Chemical Society*, *108*(8), 2040–2048.

30. Gryaznov SM, Letsinger RL. Synthesis of oligonucleotides via monomers with unprotected bases. *J Am Chem Soc*. 1991;113(15):5876-5877.

31. Sekine M, Ohkubo A, Seio K. Proton-block strategy for the synthesis of oligodeoxynucleotides without base protection, capping reaction, and P-N bond cleavage reaction. *J Org Chem*. 2003;68(14):5478-5492.

32. Dolinnaya NG, Blumenfeld M, Merenkova IN, et al. Oligonucleotide circularization by template-directed chemical ligation. *Nucleic Acids Res*. 1993;21(23):5403-5407.

33. Abhyankar MM, Hochreiter AE, Hershey J, et al. Characterization of an Entamoeba histolytica high-mobility-group box protein induced during intestinal infection. *Eukaryot Cell*. 2008;7(9):1565-1572.

34. Stros M. DNA bending by the chromosomal protein HMG1 and its high mobility group box domains. Effect of flanking sequences. *J Biol Chem*. 1998;273(17):10355-10361.

35. Molecular Biology: Structure and Dynamics of Genomes and Proteomes. Consulted: 26/06/2016

36. Kilzer JM, Stracker T, Beitzel B, Meek K, Weitzman M, Bushman FD. Roles of host cell factors in circularization of retroviral DNA. *Virology*. 2003;314(1):460-467.

37. Robyt, John F, White, Bernard J. Biochemical Techniques Theory and Practice. *Waveland Press*.1990.

38. Tom M, E. F. Fritsch, Joseph S. Chapter 5, protocol 1. *Molecular Cloning - A Laboratory Manual 1* (3rd ed.): 5.2–5.3.

39. Lewis M. Agarose gel electrophoresis (basic method). *Biological Protocols*.

40. Mikkelsen SR, Cortón E. Bioanalytical Chemistry. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2004.

41. Bio-Rad Official Webpage| Application technologies | 2nd dimension electrophoresis (http://www.bio-rad.com/en-mx/applications-technologies/2-d-electrophoresis). Consulted: 26/06/16

42. Berth M, Moser FM, Kolbe M, Bernhardt J. The state of the art in the analysis of two-dimensional gel electrophoresis images. *Appl Microbiol Biotechnol. 2007;76(6):1223-1243.*

43. Mirkin SM. DNA Topology: Fundamentals. In: *Encyclopedia of Life Sciences*. Chichester, UK: John Wiley & Sons, Ltd; 2001.

44. Dickerson R. Definitions and nomenclature of nucleic acid structure components. *Nucleic Acids Res*. 1989;17(5):1797-1803.

45. Dickerson RE, Chiu TK. Helix bending as a factor in protein/DNA recognition. *Biopolymers*. 1997;44(4):361-403.

46. JENA library of Biological Macromolecules (http://jenalib.fli-leibniz.de/Piet/help/helixparameter.html) Consulted: 30/06/16

47. Vella F. Nucleic Acids in Chemistry and Biology. *Biochem Educ*. 1991;19(2):97-98.

48. Lavery R, Sklenar H. The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J Biomol Struct Dyn*. 1988;6(1):63-91.

49. Stofer E, Lavery R. Measuring the geometry of DNA grooves. *Biopolymers*. 1994;34(3):337-346.

50. Yun E, Kahn JD. Detection of localized DNA flexibility. *Nat Int Wkly J Sci ISSN 0028-0836*, Vol 368, Nº 6467, 1994, 163-166. 1994;368(6467):163-166.

51. Travers AA. The structural basis of DNA flexibility. *Philos Trans A Math Phys Eng Sci*. 2004;362(1820):1423-1438.

52. Travers A, Shore D, Baldwin RL, et al. DNA dynamics: bubble “n” flip for DNA cyclisation? *Curr Biol*. 2005;15(10):R377-R379.

53. Ball P. Focus: DNA Flexibility at the Atomic Scale. *Physics (College Park Md)*. 2012;5:134.

54. Hagerman PJ. Flexibility of DNA. *Annu Rev Biophys Biophys Chem*. 1988;17:265-286.

55. Dickerson RE. DNA bending: the prevalence of kinkiness and the virtues of normality. *Nucleic Acids Res*. 1998;26(8):1906-1926.

56. Geggier S, Vologodskii A. Sequence dependence of DNA bending rigidity. *Proc Natl Acad Sci*. 2010;107(35):15421-15426.

57. Dlakic M, Harrington RE. Bending and torsional flexibility of G/C-rich sequences as determined by cyclization assays. *J Biol Chem*. 1995;270(50):29945-29952.

58. Thompson JO. HOOKE’S LAW. *Science*. 1926;64(1656):298-299..

59. Håkan Markström Research Manager, Lorentzen & Wettre. Bending stiffness. 2014. Non-published

60. Zhou Y, Chirikjian GS. Conformational Statistics of Semi-Flexible Macromolecular Chains with Internal Joints. *Macromolecules*. 2006;39(5):1950-1960.

61. Kepper N, Foethke D, Stehr R, Wedemann G, Rippe K. Nucleosome geometry and internucleosomal interactions control the chromatin fiber conformation. *Biophys J*. 2008;95(8):3692-3705.

62. Dimov IT, Tonev OI. Monte Carlo algorithms: performance analysis for some computer architectures. *J Comput Appl Math*. 1993;48(3):253-277.

63**.** Binder K, Heermann DW. Theoretical Foundations of the Monte Carlo Method and Its Applications in Statistical Physics. In; 2010:5-67.

64. Fermi, Enrique; Richtmyer, Robert, D. Note on census-taking in Monte Carlo calculations.1948*. Los Alamos Archive*

65. Karp RM. An introduction to randomized algorithms. *Discret Appl Math*. 1991;34(1-3):165-201

66. Binder K, Heermann DW. Monte Carlo Simulation in Statistical Physics. Vol 0. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010.

67. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of State Calculations by Fast Computing Machines. *J Chem Phys*. 1953;21(6):1087.

68. Klenin K V., Vologodskii A V., Anshelevich V V., Dykhne AM, Frank-Kamenetskii MD. Computer simulation of DNA supercoiling. *J Mol Biol*. 1991;217(3):413-419.