ENZYMATIC TOOLS FOR STUDYING THE EPIGENOME

by

KRYSTIAN BENEDYKT UBYCH

UNIVERSITY OF BIRMINGHAM

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

School of Chemistry

College of Engineering and Physical Sciences

University of Birmingham

September 2020

Abstract

In cells, methyltransferase enzymes catalyse the methylation reaction of a numerous of biomolecules such as DNA, RNA, proteins and small molecules. Most methyltransferases require the presence and binding of a cofactor. These small molecules are methyl group donors of which the most common in nature is *S*-adenosyl-L-methionine (AdoMet). DNA methylation plays a crucial role in epigenetic gene regulation and protein function modulation. Abnormal levels of CG methylation are found in cancer cells. Hence, methyltransferase enzymes are potential therapeutic targets and their inhibitors are of a high demand. Synthetic analogues of *S*-adenosyl-L-homocysteine (AdoHcy), including N^6-modified derivatives, show promise as inhibitors of AdoMet/AdoHcy dependent enzymes. However, current synthetic methods of N^6-functionalisation is limited to inert modifications or require protecting groups in the process.

Here we present an efficient and protection-group-free synthesis of N^6-substituted AdoHcy/AdoMet analogues. A broad range of functional groups was introduced including carboxyl, amine, alkyne and azide groups. We also show that such modified AdoHcy analogues can be used in the synthesis of AdoMet analogues which are recognised by two tested methyltransferases (M.TaqI and M.MpeI).

In cancer, hypermethylation is found in regions of the genome rich in cytosine and guanine, so-called CpG islands; yet, across the entire genome hypomethylation is the dominant trait. Even though the methylated cytosine residues can be detected with single nucleotide resolution using sequencing methods, little is known about the impact of DNA methylation in chromatin organization in nucleus.

Herein, we report a novel approach for the three-dimensional visualisation the architecture of accessible unmethylated DNA in mammalian cells. We employ methyltransferase-directed labelling of the epigenome using synthetic cofactor analogues for these enzymes bearing a clickable moiety for fluorescent labelling of DNA We combine this with expansion microscopy and inverted selective plane illumination microscopy (iSPIM) to reveal the organized structures of unmethylated regions of chromatin in healthy and cancerous cells.

Acknowledgements

I would like to thank all who accompanied me throughout my PhD journey and provided me with a never-ending support, guidance and help.

First and foremost, I would like to thank my main supervisor, Dr Robert Neely for giving me the opportunity to be a part of such multidisciplinary project and a chance to explore many exciting ideas, patience with reading and correcting this piece of work, the help and guidance he has given me throughout. I would also like to thank my second supervisor, Dr Francisco Fernandez-Trillo for a chance to be a part of his group in which I widened my knowledge on polymers and electrochemistry, for his help, guidance and challenges in the synthetic part of the project. I would like to thank Dr Clare Davies for giving me an opportunity to be a part of her group and explore the world of cell biology, and the help, support and guidance throughout.

Next, I would like to thank all past and present member from all three groups I worked with They helped me offering me a countless of advices, took part in endless discussions during which many exciting ideas emerged, encouraged me and supported in high and lows of my project. Especially I would like to thank Andrew Wilkinson for introducing me to the chemistry of AdoMet at the beginning of my journey, his guidance and experience; Darren Smith for his MATLAB knowledge which made my life easier many times, agreeing to be a part of never-ending discussions about life (which were mostly monologues); Ashleigh Rushton for being very welcoming during my first months in a new group, teaching me how to work properly with bacteria and being a great gym buddy after long days in the lab; Emma Faulkner for showing me secrets of microscopy and expansion microscopy; James Jarrold for all the help, advice

and patience during my exploration of biology; Jenna Jafri for a fascinating discussions and a massive help with English in my thesis; Elodie Jagu for being a tremendous support during my project, an amazing friend and companion in expeditions to Spar, struggles at the gym and a great time with wine and cheese.

Finally I would like to thank my friends and family for supporting me along the way. Special thanks go to people that made my last four years in a new city and a new country much more easier and enjoyable: Adam and Aleksandra for inviting us four years ago which was a beginning of many trips around the UK, games evenings or simple coffee, cake and chill time; my fiancé who accompanied me and supported along the way shared with me all my highs and lows during this fascinating journey.

Table of Contents

Table of Figures

I. Introduction

I.1. Epigenome

The genetic code stores information about proteins and RNA in all living organisms and is encoded within DNA molecule.[1] Over the last few decades researchers were eager to find complete genomic DNA sequences, which were considered to provide an ultimate knowledge about the function of cells and whole organisms. Especially sequence of the human genome was predicted to give answers about human evolution or the source of diseases.[2] Even though the human genome has been almost fully sequenced, as a part of an international research project - Human Genome Project, a lot of questions still remain. One of them is '… how genes orchestrate the construction and maintenance of the miraculous mechanism of our bodies…'.[2] To answer this question a number of international projects have started over the last years e.g. Human Epigenome Project which aims to decipher the epigenetic code.[3]

An epigenome is a set of the chemical modifications to DNA and histone proteins - building blocks of chromatin - that modulates their function in cells.[4] The epigenome is a key factor of gene regulation[5], tissue differentiation and development.[6] It has been found that epigenetics plays a crucial role in various diseases in human such as neurological disorders i.e. Alzheimer's,[7] autoimmune diseases i.e. lupus,[8] cardiovascular diseases[9] or various types of cancer.[5, 10, 11]

I.1.1. Epigenetic marks

I.1.1.1. DNA methylation

DNA is primarily a store of information. It is a polymeric molecule, and ordinarily has a double-stranded helical structure. The monomeric units of DNA strands are four

different nucleotides bearing heterocyclic nucleobases (adenine (A), guanine (G), cytosine (C) and thymine (T)) attached to a five member ring sugar – deoxyribose - via a glycosidic bond. The sugar contains a phosphate group attached to 5' carbon. The strands of the double helix are bound together by hydrogen bonds between complementary nucleobases following base pairing rules – A with T and C with G (Fig. 1 A).[12] DNA as a double helix can be found in three different conformations in nature: A-DNA, B-DNA and Z-DNA, with B-DNA being a most abundant structure in cells.[13] B-DNA is a right-handed helix with a complete turn occurring every 10 base pairs with a length of each turn of 34 Å (Fig. 1 B). In a structure of B-DNA two types of twists can be identified: a narrower minor groove and a wider major groove can be identified. Due to their structural difference, both grooves can be recognised differentially by various proteins and can play a role in regulation of cellular processes such as gene expression.[14]

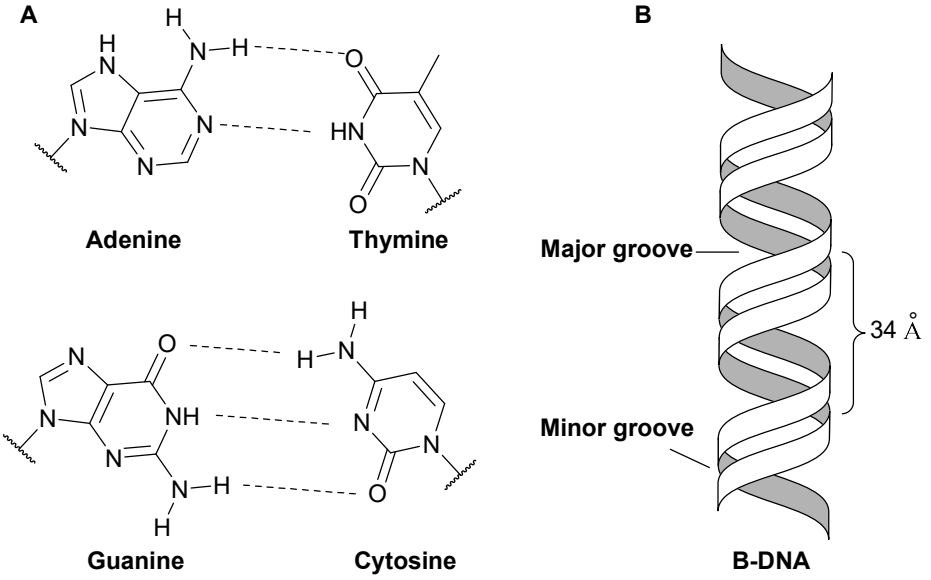


Fig. 1 (**A**) Base pairing in DNA (A-T and G-C). (**B**) a simplified structure of the most common conformation of double DNA helix - B-DNA.

Studies show that other DNA conformational structures such as Z-DNA due to their unique left-handed helical structure may be recognized by specific proteins and therefore play a role in regulation of cellular processes e.g. transcription.[15] The DNA encodes genetic information within its sequence which is utilized during gene expression. Firstly, during transcription, DNA is a template for a newly synthesized RNA strand with a complementary sequence. The product RNA can be either functional RNA molecules e.g. transfer RNA (tRNA) or protein encoding RNA – messenger RNA (mRNA). Subsequently, a protein encoded in mRNA is synthesized via a process known as translation.[16] The regulation of gene expression is a complex process and is essential for e.g. cell differentiation and metabolism.[17] As an example, neurons, skin or heart cells contain the same genomic DNA, the only difference is distinct sets of genes being active in each of them.[18]

Chemical modifications to DNA is one mechanism used in cells in the regulation of gene expression at the transcriptional level. The most common epigenetic mark of DNA is methylation. The concept of DNA methylation as a regulator of gene expression has been known for decades. In 1975 Riggs, Holliday and Pugh suggested that DNA methylation is responsible for X chromosome inactivation in mice.[19]

In bacteria, the modification of DNA is employed to distinguish host DNA from the DNA of invading viruses or phage. Methylation plays an important role in the defence system of bacteria. Moreover, methylation is a part of protection mechanism against restriction enzymes degrading DNA as well as a methyl-directed repair system of a mismatch in DNA sequence.[20, 21] During the replication process of methylated DNA, the newly synthesized strand of DNA is unmethylated. This results in a hemimethylated product with a methylated template sequence. If an error is made during the replication process

such as insertion, deletion or mis-incorporation of a nucleobase, a repair system is able to distinguish an incorrectly synthesized strand from a template strand using the methylation status. In eukaryotes, the methylation process plays a role in the regulation of gene expression. Therefore, it is crucial in processes such as DNA transcription, early development and gene silencing where it impacts the interactions of DNA with proteins involved in these processes.[22]

Two of four DNA nucleobases can be methylated via a reaction catalysed by DNA methyltransferase enzymes (MTases). Carbon 5 and C4 amino group in cytosine (m5C and m4C respectively) and C6 amino group in adenine (m6A) can be methylated (Fig. 2).

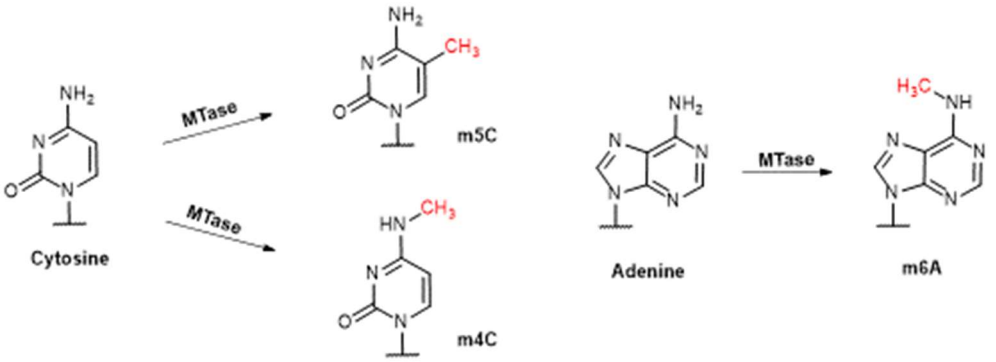


Fig. 2 A scheme of the methylation of cytosine and adenine in DNA with methyltransferase enzymes.

All of these modifications are common in prokaryotes,[23] whereas in eukaryotes the most abundant is m5C with the vast majority of these being present within the CpG dinucleotide sequence.[24] In mammals, most of the time 70% to 80% of all cytosine in CpG sequence are methylated.[25]

As the m5C especially in the CpG dinucleotide sequence plays a regulatory role, this modification can be dynamically introduced and erased in order to maintain variable gene expression levels. In mammals the m5C mark can be erased by a multistep

reaction.[26] In 2009, enzymes catalysing the m5C active demethylation process were discovered.[27] Firstly, the methyl group undergoes successive oxidation reactions to 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC) 5-carboxylcytosine (caC) catalysed by TET (ten eleven translocation) proteins.[28] In the following step, fC or caC can be removed by thymine DNA glycosylase (TDG) and further a new cytosine is restored in DNA repair process (Fig. 3 A).[26, 29] Another identified demethylation mechanism is via deamination of m5C which yields a thymine which is subsequently replaced during the DNA repair process (Fig. 3 B). The methylation mark can also be removed via passive processes, such as replication.[30] In chromatin, modifications to DNA regulate cellular processes alongside modifications to another building block of chromatin – histones.

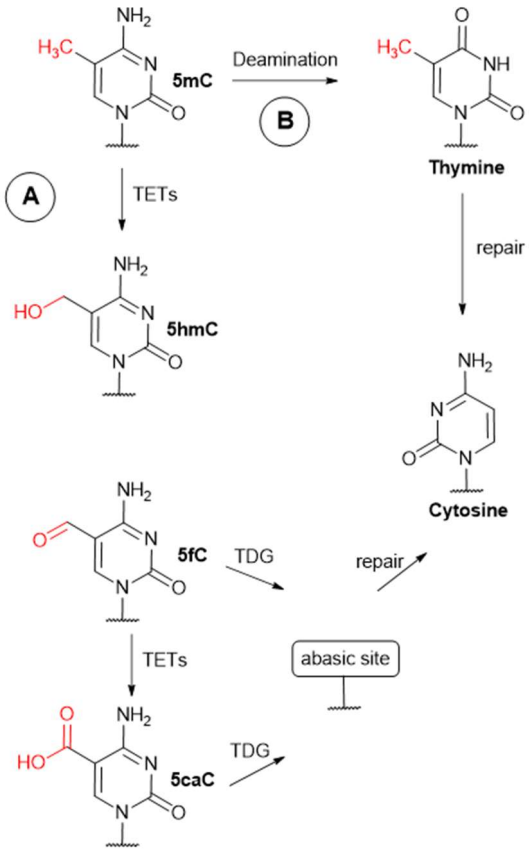


Fig. 3 Reactions scheme of m5C demethylation via a TETs mediated demethylation process (**A**) and a deamination of cytosine (**B**).

I.1.1.2. Histone modifications

Genetic code written in the DNA sequence contains all information required for living cells and organisms. A significant factor governing the accessibility of the parts of the genetic code and, therefore, expression levels are regulated by histone marks. The importance of histone marks in upstream regulation of proteins synthesis, recruitment of various proteins via specific interactions and extending the DNA code led to the hypothesis of 'histone code'.[31]

Histones are proteins located in nuclei of eukaryotic cells that form, with DNA strands, the elementary building blocks of chromatin called nucleosomes. There are four families of core histone proteins: H2A, H2B, H3 and H4 and one H1/H5 so called the linker histones.[32] In the structure of all core histones, the histone fold domain is present. It consists of three alpha helices separated by two loops that are important in nucleosome formation.[33] In addition, all histones contain tails, less than 40 amino acids long that are rich in such residues as lysine (K), arginine (R), serine (S) or threonine (T) which can undergo numerous posttranslational modifications.[32] Amongst various chemical modifications of histone tails are: methylation (lysine, arginine), acetylation (lysine), phosphorylation (serine, threonine, tyrosine) and ubiquitination (lysine).[34, 35] All the modifications are significant factors in many processes in cells. Even before DNA methylation studies, the role of histones as repressors in gene expression was anticipated by Stedman and Stedman in 1950.[19] Methylation and acetylation play a crucial role in gene expression via regulation of transcription,[36] phosphorylation in chromosome condensation during mitosis and meiosis, DNA damage repair and transcription regulation.[37]

Lysine and arginine can be methylated more than once unlike acetylation which occurs only once at the same residue. These reactions is facilitated by arginine or lysine-specific protein methyltransferases (PRMT and KMT respectively). Lysine can be mono-, di- or trimethylated whereas arginine mono- or dimethylated (Fig. 4).[31]

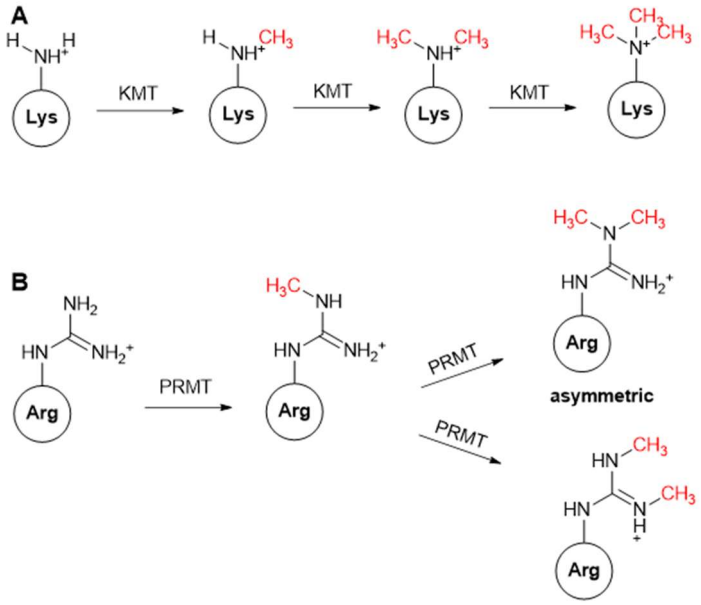


Fig. 4 MTase mediated methylation of lysine (**A**) and arginine (**B**).

However, two different types of dimethylated arginine exist symmetrical and asymmetrical that can show drastically different properties e.g. symmetrical methylation of H4R2 is an activating mark whereas asymmetrical methylation of H4R2 is a repressive mark.[38]

The effects of some of the modifications to the histone tails residues in transcriptional regulation are presented in the table below (code for the position in histone is e.g. H3K4 – histone 3 lysine 4, modifications are labelled – mono-methylation (me1), di-methylation (me2), tri-methylation (me3), acetylation (ac)).[39-42]

Modification	H3K4	H3K9	H3K27
me1	activation	activation	activation
me2	activation	repression	repression
me3	activation	repression	repression
ac	activation	activation	activation

I.1.2. The role of epigenetics in the regulation of cell processes

I.1.2.1. Epigenetics as a regulation factor of gene expression

Gene expression in cells is process of the synthesis of functional product e.g. protein, encoded in DNA sequence. In eukaryotic cells, gene expression takes place in the nucleus and cytoplasm and is a result of a cascade of following processes:

- transcription – synthesis of complementary RNA
- RNA splicing – processing of newly synthesized RNA
- translation – protein synthesis from RNA
- post-translational modifications – 'protein maturation' by covalent modifications e.g. phosphorylation

Transcription is carried out by RNA polymerase that, together with various transcription factors, binds to the promoter region of the gene to be transcribed. Two strands of DNA are locally separated and the enzyme moves 3' → 5' along the DNA strand whilst adding complementary ribonucleotides to a newly synthesized RNA strand. If the transcribed DNA fragment encodes for a protein, messenger RNA (mRNA) is created. Alternatively, DNA can encode for non-coding RNA. These RNA molecules can be employed in different aspects of gene expression and its regulation. Examples

of such non-coding RNA includes transfer RNA (tRNA), ribosomal RNA (rRNA), microRNA or ribozymes which are RNA molecules with enzymatic properties. In addition to the production of biomolecules, DNA also contains regulatory sequences such as promoters, enhancers or silencers, usually close to expressed genes, that are responsible for the regulation of gene expression. These regions are rich in specific sequences that are recognized by transcription factors that can act as activators or repressors of transcription and chromatin remodelling enzymes.[43] It has been shown that these motifs can be used to predict epigenetic modifications e.g. G+C rich regions correlate with repressing modifications such as H3K27me3 and H3K4me3. On the other hand many sequences recognized by transcription factors were found to be co-located with H3K27ac marks. Acetylation of lysine 27 is enriched in active promoters and enhancers, where chromatin is very dynamic, suggesting an important role for those motifs in the establishment of epigenetic marks.[44]

Together with a set of histone marks, methylation of DNA plays a crucial role in regulation of transcription. The majority of promoters contain CpG islands with high densities of CpG dinucleotides i.e. usually over 200 bp long sequences with an increased content of G+C bases (>50 %) and a ratio of observed to expected number of CpG sequence greater than 0.6.[45] In general, methylation of the CpG sites causes the gene associated with the methylated promoter to be silenced.[46] On the other hand, m5C is enriched in the body of highly transcribed genes, which would suggest its opposite effect to gene silencing in promoters.[47]

In general dysregulated methylation of DNA may lead to disruption of cell-cell interaction, disruption of cell signalling or genetic instability. In the development of cancer, hypermethylation of so called tumour suppressor genes, thus inhibition of the

synthesis of proteins critical for regulation of cell cycle, DNA repair or transcription. This can lead to the overgrowth of cells or DNA damage.[48] The mechanism of regulation of gene transcription by DNA methylation is complex and progress in understanding has been made over last years. A family of methylated-CpG binding domain proteins (MBDs), that inhibits gene expression by recognition and binding to hypermethylated regions, has been identified. MeCP2 and MBD2 are known for a recruitment of co-repressors of transcription, chromatin remodelling proteins and histone deacetylases and preventing the binding of transcription factors to DNA.[49, 50] It was thought that DNA methylation itself can prevent transcription factors from binding to DNA and therefore inhibit transcription. However, recent studies suggest that DNA methylation can modify recognition by transcription factors. which may be able to recognise and bind to different sequences depending on their methylation status. Alternatively, methylated cytosines within gene bodies correlate positively with highly transcribed genes.[51] These regions are usually poor in the CpG sequence but enriched in repetitive sequences including retroviral repeats and transposable elements. Although the function of this modification is not clear it is thought that gene body methylation silences repetitive DNA elements, which might be disruptive for cells, whilst enabling the elongation of the transcription.[52]

I.1.2.2. DNA methylation in repetitive regions

Repetitive sequences are widely spread across the human genome and are thought to compose up to two-thirds of the sequence of the human genome.[53] The repetitive elements can appear as copies of a few up to hundreds of nucleobases long sequences, e.g. microsatellites or encoding functional protein sequences repeated

across the genome such as SINEs (short interspersed nuclear elements) and LINEs (long interspersed nuclear elements).

SINEs and LINEs are a part of transposable elements in the genome that can change their position in the genome. This can lead to a development of new genes or regulatory elements or variation between individuals even though most of them are considered to be neutral.[54] LINEs, unlike SINEs, encode the reverse transcriptase protein that translates and integrates RNA into DNA. Therefore, upon expression, reverse transcriptase is able to write the original LINE sequence, or SINE sequence, back into the genomic sequence at a new site. As their activity could lead to uncontrollable mutations most of them are methylated in the genome and hence silenced. In cancer transposable elements are hypomethylated and their activity leads to increased insertional mutagenesis.[55]

Tandem repetitive sequences do not have the ability to change their location and are located mostly in gene poor regions such as centromeres, pericentromeres and subtelomeric regions of the chromosomes but can also be found within gene sequences or non-coding sequences such as regulatory elements. They are highly unstable and are prone to mutations such as insertions or deletions of usually an entire repeat sequence, hence their importance from an evolutionary point.[56] However, tandem repeats in both coding regions and regulatory elements such as promoters or enhancers can affect gene expression. In such regions transcription factor binding sites, usually short sequences of DNA recognised by transcription factors, occur usually in clusters. A presence of unstable tandem repeats in such regions can change the number of transcription factor binding sites subsequently influencing transcription of respective gene. As such, in normal cells tandem repeats are mostly methylated

whereas in cancer, similarly to transposable elements, tandem repeats are hypomethylated.[57]

Chromatin interacts with a plethora of proteins including chromatin remodellers or transcription factors. Some of them are sensitive to modifications of DNA and histones which, hence these marks can affect the structure of the chromatin in the nucleus modifying the accessibility of these parts of the chromatin to further complexes of proteins. Therefore, it is important to study how the organisation of chromatin influence cellular processes.

I.1.3. Structure and organisation of chromatin

I.1.3.1. Chromatin structure

Human DNA contains about 3 billion nucleotides which when stretched would be about 2 meters in length. This is all packed in a single cell and organized in a structure called chromatin. There are a few stages of DNA organisation that can be distinguished within chromatin in eukaryotes. The basic unit of DNA compaction is a nucleosome. It consists of approximately 146 base pairs (bp) wrapped around an octamer of histone proteins H2A, H2B, H3 and H4 (Fig. 5).[58] Histone 1 acts as an anchor, binding entering

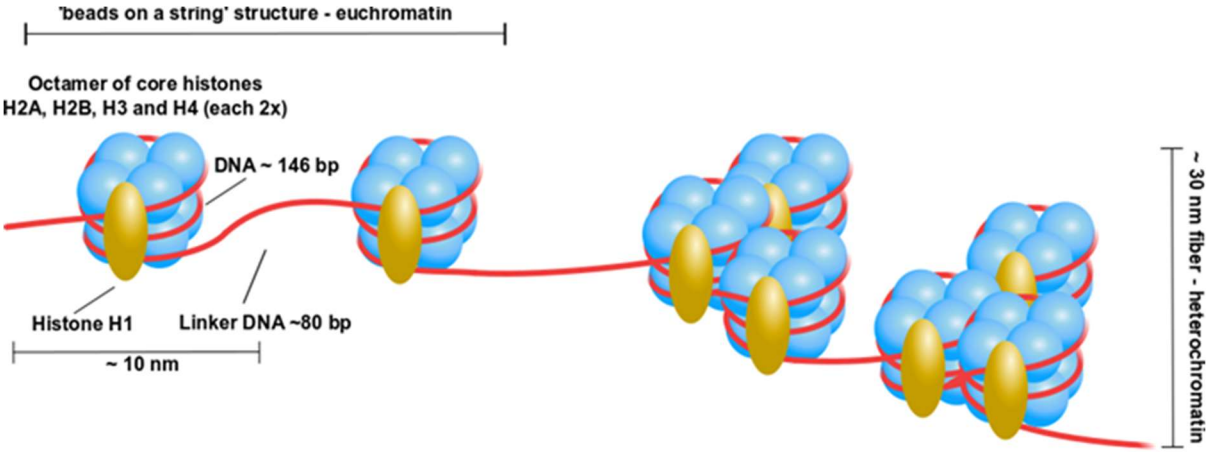


Fig. 5 Individual nucleosomes form a 'beads on a string' structure' – euchromatin - which then can condense into heterochromatin fibres.

and exiting DNA to the nucleosome. [59] Between the nucleosomes is 80 bp of linker DNA. Many nucleosomes form a structure with a 10 nm diameter, called 'beads on a string' - euchromatin.[60] Due to its relatively loose structure, euchromatin is considered as 'active' or 'poised' part of the genome, and can be readily transcribed.[31] Multiple nucleosomes can be coiled in higher order structures creating 30 nm in diameter fibres which form heterochromatin. This is a silent part of the genome due to restricted access of the transcriptional machinery to DNA strands.[31, 61]

I.1.3.2. DNA is organised in nucleus

During short mitosis division, chromatin undergoes very dynamic condensation processes forming the densely packed metaphase chromosomes. In 1885 Carl Rabl suggested that during interphase, when chromatin appears in a relaxed form, chromosomes show a spatially discrete distribution.[62] It was named by Theodor Boveri as chromosome territories (CT)[63] and later confirmed that rather than chromatin being randomly interspersed, chromosomes occupy certain positions in the cell nucleus (Fig. 6).[64, 65] Heterochromatin and euchromatin form Mbps long chromatin domains (CD) which ultimately organize in CT. [66-68] After almost 20 years, imaging of CTs in living cells showed that the position of each CT within a nucleus remained similar throughout the interphase and correlates with the features of CT such as gene density or chromosome size.[69-71] The shape of CT itself was also linked to the gene activity. It was shown that gene rich and active CTs are more irregular in shape and less compact.[72, 73] Actively transcribed and early-replicating foci have been found to occupy the interior of the nucleus, whereas silenced genes and late-replicating sites were preferentially found in the outer parts of the nucleus.[74, 75] As an example, in female cells inactive chromosome X is more regular in shape than the active one and found

in more peripheral parts of the nucleus.[76, 77] Although the morphology of CTs is generally stable, at the Mbps scale dynamic oscillations between open and condensed structures were observed within CDs and were correlated with dynamics of nuclear processes i.e. replication and transcription.[78-80] Similarly to global euchromatin and heterochromatin, the active, open chromatin and repressed part within a single CT was divided into two compartments, A and B respectively (Fig. 6).[81] Recently, it has been suggested that instead of binary A/B structures, there is a continuous spectrum of chromatin states between A and B.[82]

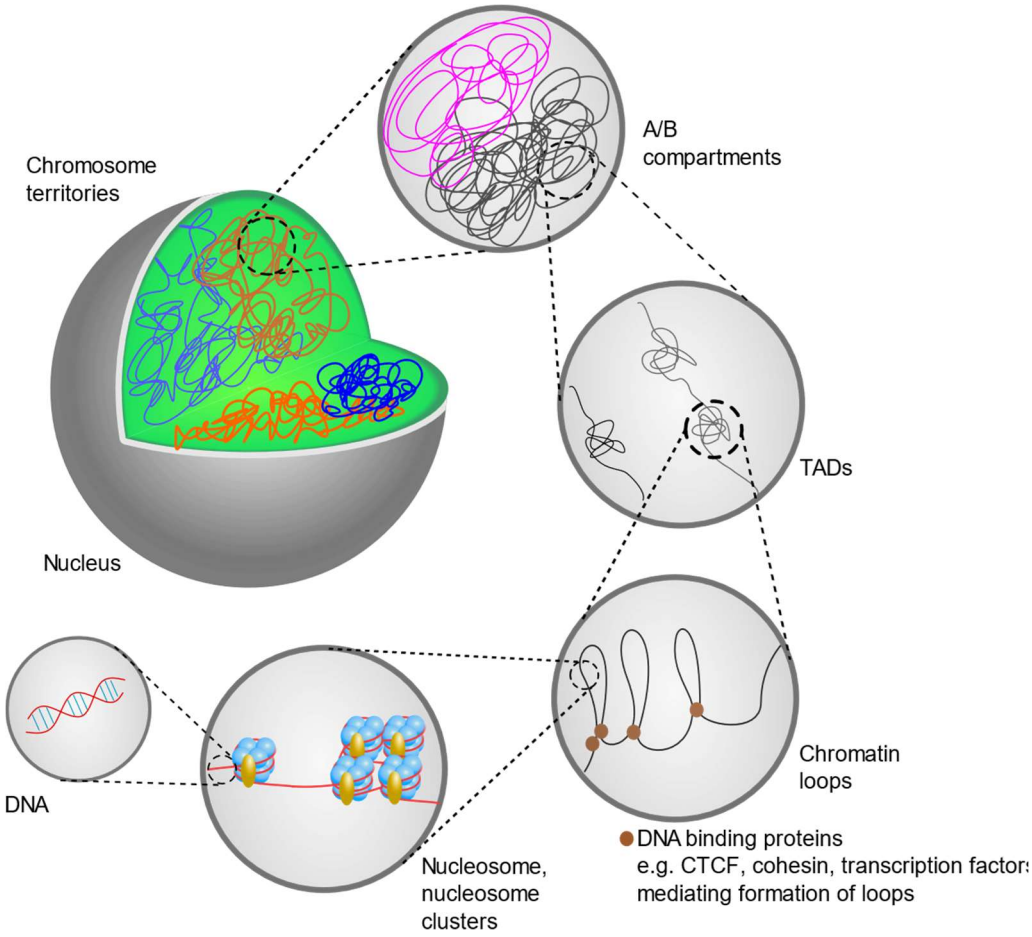


Fig. 6 A simplified structure of chromatin in the nucleus.

I.1.3.3. Function of the structure of chromatin on cellular processes

A great challenge is to understand how spatial organisation, expression and regulation of thousands of genes is associated with the three-dimensional structure of the nucleus.[62] When investigated in more details, the structure of cell nucleus shows a plethora of structures. The main processes present in the nucleus such as DNA replication, transcription and RNA splicing are spatially organized in discrete sites. Alongside with small foci where these processes occur, larger globular structures, calles nuclear bodies, containing a number of nuclear factors such as proteins and RNAs, can be found.[83] There are a variety of nuclear bodies present including Cajal body, nuclear speckles, histone loci and the nucleolus. Cajal bodies are related to splicing via processing and modification of spliceosomal small nuclear RNAs and are important in telomere maintenance. The nucleoli contain DNA, RNA and proteins and play a crucial role in the synthesis of proteins as sites of ribosome biogenesis.[84, 85] Although the mechanism of the creation of the nucleolus is unknown it has been proposed that proteins binding to DNA strands create a polymer-like structure that causes the collapse of chromatin and formation of globular structures. Alternatively, the interaction between proteins and chromatin leads to liquid-liquid phase separation in the nucleus, with nuclear bodies and chromatin forming two separate phases.[86, 87] A number of structures in the nucleus are spatially organised with the participation of the nonchromatin elements of the nucleus including the nuclear matrix.

The nuclear matrix is hypothesized as a scaffold that helps to organize and maintain the subnuclear structures without necessarily directly influencing gene expression. The nuclear matrix is considered to be a dynamic network of fibres associated with proteins and RNA, [88] although the evidence of their organisation and function *in vivo* are still

not clear.[89] It has been found that chromatin interacts with a number of nuclear matrix proteins via recognition sites located on DNA called S/MARs (scaffold or matrix attachment regions). These regions occur in repetitive sequences of the genome usually rich in A-T nucleobases and transcription binding sites.[90] A few of the proteins involved in anchoring the genome to the nuclear matrix are topoisomerase-II, CTCF, cohesin and nuclear lamins.[62] An important protein in this process is CTCF, a protein that binds to DNA that can act as a transcription activator/repressor or an insulator e.g. preventing the spreading of methylation from adjacent regions in the genome by blocking DNA methyltransferase activity (Fig. 7).[91] In breast cancer, a loss of CTCF binding leads to propagation of methylation onto a promoter of the BRCA1 cancer suppressor gene causing its inhibition and hence cancer propagation. In addition to transcription regulation it plays an important role in gene imprinting and recombination. Moreover, a loss of CTCF on specific loci leads to a methylation of an active allele (Fig. 7), which is found in colorectal cancer and chronic myeloid leukaemia.

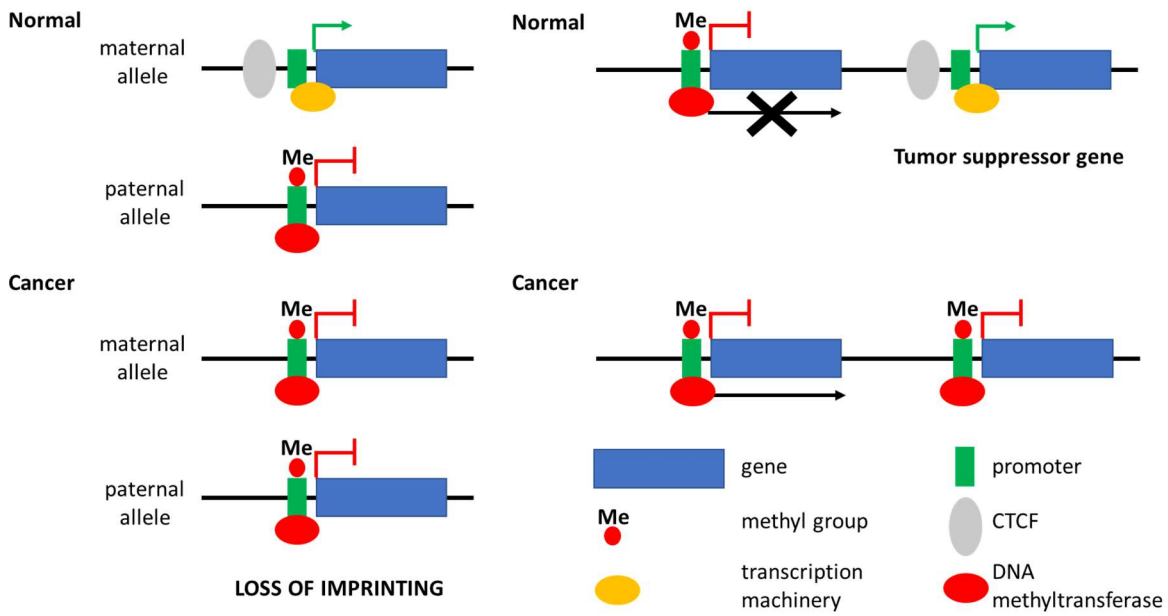


Fig. 7 Role of CTCF in gene imprinting (left) and DNA methylation mediated gene expression (right).

Through the mediation of chromatin loops formation, CTCF and cohesin bring together distinct regions of DNA enabling their correct interaction, e.g. promoter and enhancers in gene expression (Fig. 6).[91-93] These clusters of interacting chromatin loops are called topologically associated domains (TADs).[94, 95] CTCF regulates TAD boundaries by occupying DNA fragments linking two different TADs and therefore acting as an insulator. TAD boundaries are also enriched in housekeeping genes, and tRNAs.[96] The analysis of recent chromatin conformation capture methods (e.g. 3C, Hi-C) suggest that TADs are equivalent to chromosome domains (CD).[97]

An important part of nuclear matrix is the nuclear lamina also referred to as the nuclear envelope. It consists of various proteins with a major three being lamins A, B and C. Nuclear lamina is a key factor in nuclear organisation as well as regulation of many processes via interaction with proteins such as histones, actin or transcription and replication complexes.[98] Clusters of chromatin loops, usually low in gene density and rich in repetitive elements, anchored or/and interacting with nuclear lamina were named as lamina-associated domains (LADs).[99, 100] Although impact of the nuclear lamina on the induction of repressed LADs is still unknown the explanation was suggested that a contact of parts of chromatin with the nuclear lamina promotes the activity of enzymes responsible for chromatin repression such as histone deacetylase. Therefore, the nuclear lamina plays a role in chromatin organisation, recruitment of epigenetic regulators and are associated with gene repression. Similarly to LADs, parts of heterochromatin associated and interacting with nucleolus were identified and termed nucleolar-associating domains (NADs). These regions are enriched in inactive ribosome coding DNA which loops into nucleolus when transcribed.[101]

In addition to the nucleolus, DNA undergoes transcription across the nucleus. It has been observed that transcription can be spatially localised in foci called transcription factories.[102-104] Such a local concentration of the transcriptional machinery enables efficient gene expression where multiple genes can be transcribed in one factory.[104] The size of transcription factories varies across the cell cycle, with an average of 87 nm in *HeLa* cells.[105] Transcription factories consist of protein-rich core regions surrounded by densely packed chromatin.[106] The formation of loops of chromatin is an important mechanism of bringing distant loci of the same or separate CTs to the same locus, hence chromatin organization is important for the efficiency of transcription.[107] Transcription factories were found to be enriched in promotor-promotor and promotor-enhancer interactions. A key role in the localisation of promoters, enhancers and other regulatory factors play chromatin remodelling complexes and insulators such as aforementioned CTCF. Even though there is a number of studies supporting the existence of transcription factories, some recent studies suggest that RNA polymerase II protein, involved in transcription, may cluster only temporarily rather than form metastable centres of transcription.[108]

I.1.3.4. Epigenetics as a regulator of chromatin organisation

DNA methylation is a key factor in chromatin organisation. On a scale of a single DNA strand, a methylation of cytosine residue promotes a conformational change from B- to Z-DNA form. On a larger scale DNA methylation can impact on the binding of DNA to many proteins. As an example methylated DNA strands wrap more tightly onto the histone octamer leading to formation of more stable and compact nucleosomes.[109, 110] DNA methylation regulates the function of many proteins and binding factors e.g. methylation of the CpG sequence within recognition sequence of insulin-like

growth factor 2 is important in the process of gene imprinting. It hinders CTCF binding which enables passing the methylation status of a gene after cell division which would be otherwise blocked by CTCF bound to unmethylated region (Fig. 7).[111] Methylation regulated CTCF binding influences the structure and organisation of TADs. It was shown that hypermethylation of CTCF binding sites leads to repression of insulator function of CTCF resulting in abnormal interactions between regulators and genes from separate TADs e.g. activation of oncogenes in cancer.[112] An impact of DNA methylation on chromatin organisation is also associated with transcription. The nucleolar regions of the nucleus show highest transcription rates in the nucleus, as ribosomal DNA (rDNA) needs to be processed for the synthesis of new ribosomes and thus enable cell amplification of the genomic information and eventually proliferation. It has been found that the rDNA gene body is rich in CpG dinucleotide sequences and the hypomethylation of rDNA promoters, regulated by the NuRD complex (nucelosome remodelling and and histone deacetylase), keeps them poised for transcription.[113] However, induced hypomethylation of CpG-sequences by either a loss of DNMT1 and DNMT3b methyltransferases or synthetic inhibitor of methyltransferases (azacitidine) caused a deformation of nucleoli, loss of compaction, increased number of nucleolar foci (Fig. 8) and dysregulation of rRNA synthesis.[114, 115]

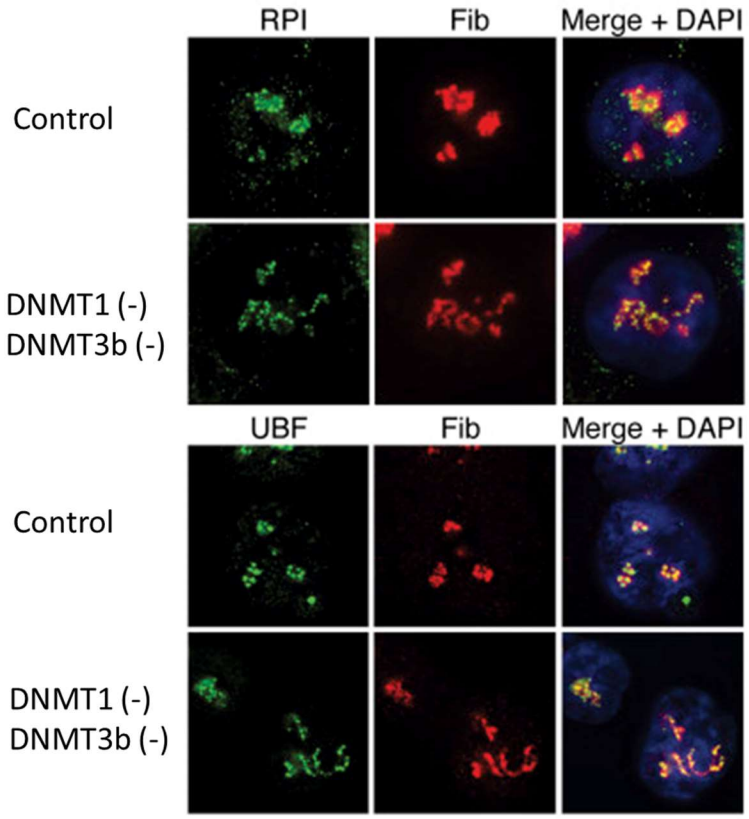


Fig. 8 Inactivation of DNMTs leads to an increase in RPI (RNA polymerase I) and UBF (transcription factor)-Positive Gene Foci and to nucleolar dispersion. Fib – fibrillarin – a nucleolar protein. Figure from Gagnon-Kugler T., *Molecular Cell*, 2009.

In addition to DNA methylation histone modifications impact the structure of the chromatin. Repressive histone marks such as H3K27me3 found in LADs, induce the activity of chromatin remodelling proteins promoting the formation of heterochromatic structure of chromatin structures. [100] Some histone modifications are important for the anchoring of LADs to the nuclear lamina. The depletion of activity of proteins introducing H3K9me2/3 marks decreases binding of these regions to the nuclear lamina.[116]

Genome organisation changes dynamically during stem cell differentiation. It has been shown in human embryonic stem cells (hESCs) that interaction between TADs changes drastically. The increased interactions was observed between TADs with

increased level of gene expression with enrichment in active epigenetic marks, e.g. H3K4me1, and reorganisation from compartment B to A.[117]

I.1.3.5. Aberrant chromatin structure in diseases

Proper regulation of three-dimensional genome architecture is important in maintaining normal cell functions. Aberrant changes, e.g. single nucleotide polymorphism (SNP), in noncoding regions of DNA located in or near regulatory elements, such as promoters, enhancer or silencers that interact with chromatin remodelling proteins or transcription factors, affect transcription of specific genes and can cause various diseases. Such changes to genome sequence can regulate activity of very distant genes by a creation of chromatin loops, and, hence, the three-dimensional organisation of chromatin is a key element of this process.[94] As shown using DNase I hypersensitivity method, due to increased level of SNPs these regions change their 3D structure to more open form which then can be recognized by various transcription factors or close form which hinders this interactions.[118] Changes in the structure of these regions were associated with various diseases such as multiple sclerosis, Parkinson's disease, diabetes, Rheumatoid arthritis and breast cancer.[118]

Abnormal levels of DNA methylation in cancer are characterized by site-specific hypermethylation of CpG islands, especially at promoters, in contrast to genome-wide hypomethylation, e.g. in repetitive regions.[5, 119] Hypermethylation of CpG islands correlates with hypoacetylation of histones leading to reorganisation of chromatin structure into condensed and hence supressed state. As a result, genes responsible for cell-cycle regulation,[120-122] chromatin remodelling,[123] transcription,[124] DNA repair,[125] apoptosis[126] etc. become silenced leading to further changes in chromatin organisation and cancer development. On the contrary the abnormal hypomethylation of these

regions causes genome instability due to increased mitotic recombination and transposon activity.[119] Therefore, various proteins associated with DNA methylation and histone modifications have been identified as potential therapeutic targets.

I.2. Methyltransferase directed labelling of DNA

I.2.1. Methyltransferases and AdoMet-dependent methylation

Methyltransferases (MTases) are a large group of enzymes responsible for methylation of a wide range of biomolecules in living cells. The methyl group is added to S, C, N, or O atoms of molecules such as nucleic acids, proteins, neurotransmitters or hormones. Due to the wide variety of natural substrates, MTases play a crucial role in biosynthesis, epigenetic gene regulation, protein function modulation, signal transduction, degradation of metabolites, etc. [127-131]

Methyltransferases fall into different classes based on types of the substrates catalysed, on the atom subjected to methylation or on the cofactor requirement. In terms of substrate specificity within one class, MTases exhibit the similarity in structure and can be divided into following groups (Table 1) [132]

Target biomolecules		**Enzymes**	**Function**
Small molecules	Catechol	COMT	monoamines metabolism and its deregulation may be related with brain diseases[133]
	Glycine	GNMT	due to the ubiquity of its substrate it exhibits a numerous regulatory roles and its malfunction may cause cancer e.g. liver cancer[134]
	Histamine	HNMT	terminates histamine activity; its polymorphism is associated with e.g. Parkinson's disease[135] or asthma;[136]
Lipids		PcaA, CmaA1, CmaA2	cyclopropane formation in membrane lipids therefore it influences bacteria resistance and virulence[137]

DNA	Cytosine C5-specific (M.HhaI, M.HaeII, M.MpeI, DNMT1, DNMT3a/b) Cytosine N4-specific (M.PvuII) Adenine N6-specific (M.TaqI, M.DpnII)	DNA control, repair, modification and restriction
Proteins	PRMT (protein arginine MT) KMT (protein lysine MT)	Histone methylation, function in cellular processes including regulation of transcription; chromatin condensation[138]
RNA	mRNA ribose 2' OMT, NS5 MT	mRNA capping and processing crucial in mRNA stabilization, protection and translation;[139]
	transcription factor B (TFB)	mitochondrial RNA transcription[140]

Table 1 An overview of AdoMet-dependent methyltransferase enzymes, their biological targets and function in biology.

The DNA methyltransferases recognize short (2-8 bases long) sequences within which targeted nucleobase can be found. This enables specific modification of exact nucleobase in entire DNA sequence, protecting DNA from being modified in an uncontrolled manner. For instance the most commonly studied bacterial MTases: M.TaqI from Thermus aquaticus and M.HhaI from Haemophilus haemolyticus methylate adenine in 5'-TCG**A**-3' and cytosine in 5'-G**C**GC-3' sites respectively. Typically DNA which has been methylated by a specific methyltransferase is protected from degradation at that site by restriction endonucleases which target those sites. In human three DNA methyltransferases modifying cytosine have been identified so far – DNMT1, DNMT3a and DNMT3b. DNMT1 is a maintenance DNA methyltransferase preferentially targeting CpG sequence of hemimethylated DNA synthesized during replication.[141] On the other hand, DNMT3a and DNMT3b are thought to function in *de*

novo methylation and target unmethylated CpG sites. Some 'off-target' methylation of non-CpG sites by these enzymes is also observed.[142]

The vast majority of methyltransferases require the presence and binding of cofactor to methylate their target sequences. These small molecules are methyl group donors of which the most common in nature is *S*-adenosyl-L-methionine (AdoMet). As a result of methylation of biomolecules, AdoMet is converted into *S*-adenosylhomocysteine (AdoHcy) which is subsequently hydrolysed to adenosine and homocysteine in presence of *S*-Adenosylhomocysteine hydrolase (Fig. 9).[143]

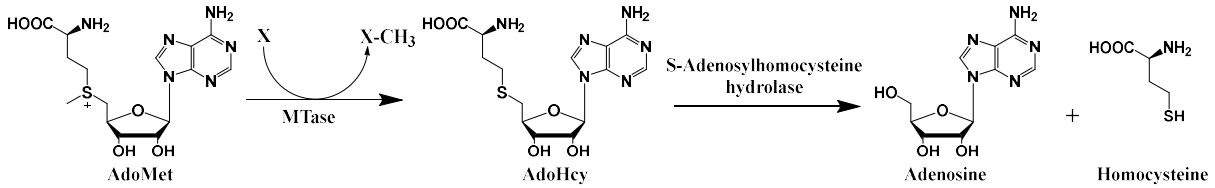


Fig. 9 AdoMet metabolism in cells. AdoMet is a donor of methyl group in methyltransferase mediated methylation of biomolecules upon which is converted to AdoHcy. In subsequent reaction AdoHcy is hydrolysed into adenosine and homocysteine by *S*-Adenosylhomocysteine hydrolase enzyme.

To complete the 'AdoMet-cycle', homocysteine undergoes methylation to methionine (catalysed by methionine synthase) which is then employed in the synthesis of AdoMet or converted into a major cellular antioxidant – glutathione.[144]

The AdoMet molecule is an unique molecule whose subunits are essential parts in many cellular mechanisms. Besides being a donor of a methyl group, AdoMet works as precursor of other biologically important molecules in further cellular pathways. In synthesis of cyclopropane fatty acids (CFAs), AdoMet serves as a source of methylene groups, in biosynthetic pathway of the biotin – necessary vitamin in cell growth – as a source of amino group. Furthermore, its ribosyl group is employed in post-transcriptional modification of tRNAs, AdoMet's aminoalkyl groups take part in the synthesis of the precursor of the plant hormone ethylene.[127]

I.2.2. Synthetic analogues of AdoMet

Due to the fact that AdoMet molecule is a very important for metabolism of living cells, its properties and interactions with different proteins have been studied for decades. One of which is binding and interaction with methyltransferases. Throughout the last decades a range of synthetic AdoMet analogues have been designed and synthesized in order to obtain desired specific interactions of novel analogues with targeted methyltransferase. The AdoMet-based cofactor analogues can be split into two groups: aziridinoadenosines and doubly-activated AdoMet analogues.

I.2.2.1. Aziridinoadenosines

Aziridinoadenosine analogues consist of an aziridine ring in place of 5'-sulfur in the structure of natural AdoMet. The synthesis of aziridine derivatives of AdoMet has been

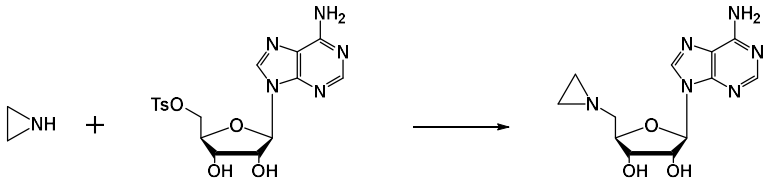


Fig. 10 Synthesis of aziridine derivative of AdoMet

reported for the first time by the group of Weinhold in 1998.[145] The synthesis is through a nucleophilic substitution of activated, with tosyl or mesyl groups, 5' hydroxyl group of adenine with aziridine (Fig. 10).

Aziridine derivative of AdoMet in order to be an alkylating agent in reaction with methyltransferases must be activated by protonation of the nitrogen atom within aziridine atom similarly to positively charged sulphur atom in natural AdoMet. Aziridine based cofactors are unique in terms of acting as an alkyl group donor. In case of AdoMet only methyl group attached to sulphur atom is incorporated into biomolecule whereas the whole aziridine cofactor analogue is an alkylation agent (Fig. 11 A).

Subsequently the 8-position of adenine of aziridine cofactors analogues was labelled with biotin,[146] fluorophores such as dansyl[146, 147] or modified with carbon chains containing reactive chemical groups e.g. azide[148] for further modification of DNA. It has also been shown that moieties can be introduced into the 6- and 7- position of adenine and the cofactor analogues remain compatible with a range of methyltransferases and transferable to their substrates.

Activated three membered rings show a high reactivity and undergo nucleophilic ring-opening reaction in presence of nucleophilic agents. This reactivity impedes the synthesis of further modified aziridine based cofactors. The synthesis of such additionally functionalized aziridine cofactors is challenging and requires several steps with low overall yield. As a result, an alternative synthetic path using an N-chloromustard as a precursor for an in situ generated aziridine ring, as a result of intramolecular cyclization of beta-halide and 5' amino group, has been described.[149] Additionally, cofactors modified with a propargyl group on the 5' nitrogen atom were obtained. Therefore, a targeted substrate e.g. DNA can be further labelled via click chemistry.

Unfortunately, the application of aziridine based cofactors has some disadvantages. Firstly, due to the product inhibition, a stoichiometric amount of MTase in transalkylation reactions is required. Secondly, the high reactivity of this class of compounds results in a risk of nonspecific alkylation and even the possibility of modification of biomolecules in the absence of a methyltransferase enzyme. Furthermore, the membrane permeability of these cofactor is not sufficient for performing studies on live cells.[150]

I.2.2.2. Doubly-activated AdoMet analogues

The second class of synthetic AdoMet analogues are doubly-activated compounds. Their methyl group adjacent to sulphur atom is replaced by a longer hydrocarbon chains usually with a functional terminal group which can be transferred onto a nucleobase of DNA by MTase (Fig. 11 B).

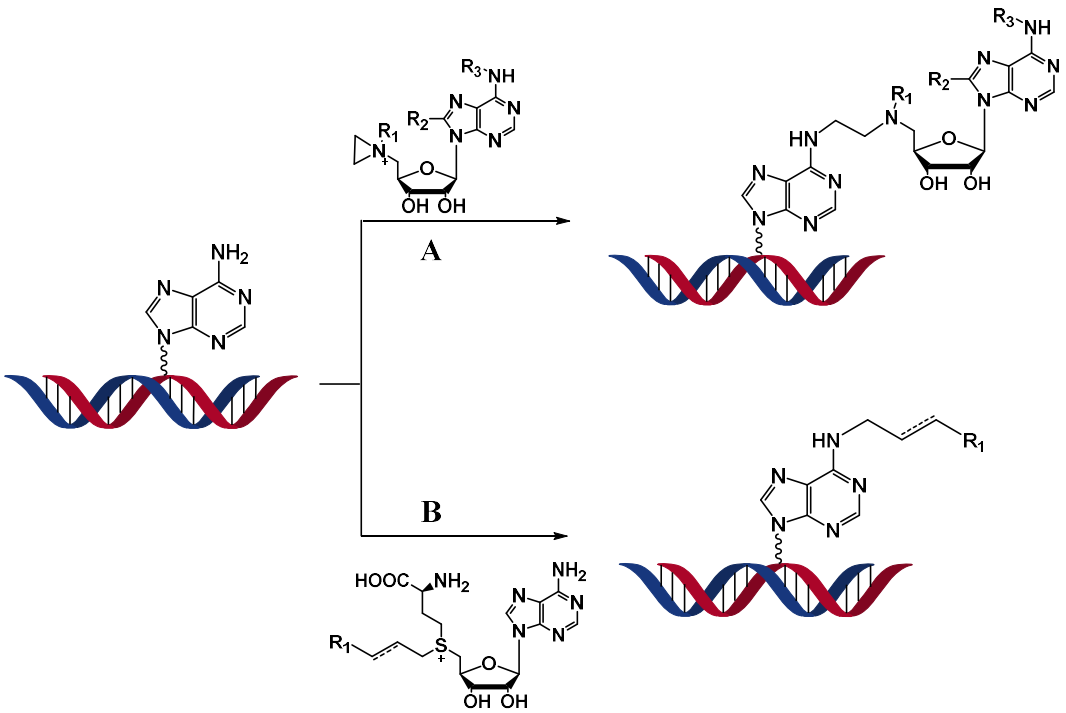


Fig. 11 Synthetic cofactors used in labelling of DNA via methyltransferase mediated reactions. (**A**) Aziridine cofactors act as a cofactor and a transferable group that can be incorporated onto DNA. (**B**) Doubly-activated AdoMet analogues bear a transferable linker that can be introduced onto a nucleobase.

It has been observed that, due to steric effects, the rate of alkylation reaction catalysed by DNA MTases massively decreases with increasing length of side chain e.g. ethyl or propyl groups. However, in 2006 Weinhold and Klimašauskas overcame this issue by introducing an unsaturated bond – double or triple – at the β position stabilizing transition state in S_N2-like transalkylation reaction (Fig. 12).[151]

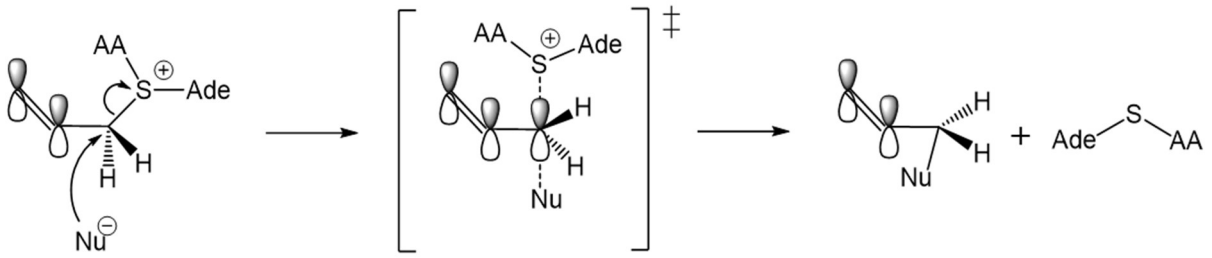


Fig. 12 A role of β-unsaturated bond in a synthetic cofactor in stabilisation of the transition state during S_N2-like alkylation reaction catalysed by methyltransferase enzymes.

The most common method of synthesis of doubly-activated AdoMet analogues is via nucleophilic substitution of sulphur center of AdoHcy with a great excess – over 50 equivalents – of strong electrophiles such as alkyl triflate or alkyl bromide under acidic conditions (Fig. 13).[151]

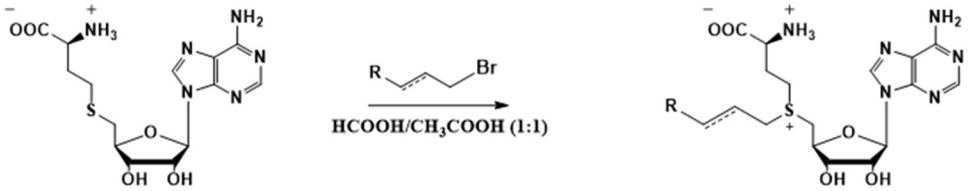


Fig. 13 Synthesis of doubly-activated AdoMet analogues via coupling of AdoHcy with alkyl bromide

These AdoMet analogues are susceptible to decomposition, which is a significant consideration in their synthesis, purification and storage. It has been reported that artificial AdoMet analogues show much less stability than natural AdoMet with half-lives from 0.05 to 5h at pH 7.4 comparing to 17h for AdoMet.[152] Usually to lower the risk of breakdown AdoMet analogues are purified and kept in slightly acidic conditions. The desired product is obtained as a mixture of two (R-) and (S-) epimers. Since, only (S)-epimer is active for transalkylation reaction, with a majority of MTases, the good separation is important. This can be usually achieved by HPLC.[150]

Another method of a small-scale synthesis of double activated AdoMet cofactors is a modified natural synthetic path of AdoMet. In cells AdoMet is synthesized from ATP and L-methionine in reaction catalysed by *S*-adenosyl transferase enzyme (MAT). This

method has been implemented to obtain a range of analogues with high percentage yields and stereochemically pure products.[153, 154]

Doubly-activated AdoMet analogues can be modified with a range of functionalized linkers containing amino,[155] azide, allyl, alkyl or acetyl groups.[156] Thus, a range of chemical strategies of labelling can be implemented to attach probes to the targeted activated biomolecules e.g. NHS strategy of active esters, azide-alkyne cycloaddition or coupling with hydroxylamine.

I.2.3. Applications of methyltransferase directed DNA labelling

I.2.3.1. Selective modification of biomolecules

In living cells, AdoMet molecule serves as a cofactor of a number of proteins and enables modification of a tremendous number of biomolecules. Hence, synthetic AdoMet analogues are versatile tools for studying cellular mechanisms in which AdoMet is involved, the function of AdoMet-dependent enzymes as well as the natural targets of those enzymes. Thanks to the selectivity of MTases in alkylation of different intracellular structures, AdoMet analogues can be used to introduce many different moieties specifically, to proteins, nucleic acids and other targets.[157]

In principle, the AdoMet analogues are recognized by multiple MTases in the cell. Hence, achieving some degree of selectivity is a key aim in this field. This can be achieved by either tuning structures of the binding pockets of methyltransferases or by synthesis of selective AdoMet analogues. In 2001, the ‘’bump-and-hole’’ approach was successfully applied to create an engineered pair of AdoMet analogue and the methyltransferase protein, PRMT1.[158] It was found that N^6-benzyl AdoMet bearing bulky group was a selective cofactor for an engineered PRMT1 mutant. This approach

may help in selective modification and studies on substrate of mutated methyltransferase in presence of other in cellular enzymatic machinery.

I.2.3.2. Genomic analysis

MTases and AdoMet based analogues can improve currently available techniques of genome analysis. Sequence specific restriction endonucleases (REases) can be utilized in DNA mapping. Thanks to its selectivity in DNA cutting, a set of REases digests a DNA sample resulting in well-defined and characteristic DNA fragments so called 'DNA fingerprint' which can be subsequently analysed by gel electrophoresis.[159] Instead of using degradation enzymes, targeted MTase-catalysed labelling can be employed. Unlike the DNA degrading restriction enzymes, using MTases enables a DNA molecule to be specifically and fluorescently labelled whilst avoiding damage to the DNA and thus increasing its overall stability. Moreover, next to more directed labelling, MTases enable high labelling densities, which increase the sensitivity of this method.[157] Due to high density of labelling and non-destructive method, linearized DNA molecules show characteristic patterns of labelling, that depend on their underlying sequence and which can be used to identify a genome or species.

I.2.3.3. Investigation of epigenetic regulation

MTases and AdoMet analogues can be useful tools in studies of DNA and protein methylation and their function in various processes that control gene activity without applying any changes in DNA sequence. As methylation plays a crucial role in these processes, MTases are ideal for identification of regions of methylation as well as unmethylated areas. As MTase can tag cytosine and adenosine, a different moiety can be introduced within each nucleobase e.g. fluorescent dyes of different wavelength of

emitted light. Hence, DNA maps and methylation maps can be generated.[157] DNA mapping based on MTases was applied in the mapping of unmethylated CpG sites which uncontrolled methylation can cause cancer.

In 2013, S. Klimašauskas and coworkers used a bacterial M.SssI MTase to introduce amine and azide functionalities to cytosine residues in the CpG dinucleotide sequence of genomic DNA *in vitro* (Fig. 14).[160] To such modified DNA, a biotin tag was introduced via click reaction followed by capture onto streptavidin beads. Purified DNA was then released from the beads and sequenced. Commonly used methods for studies of DNA methylation use antibodies against m5C which are biased towards heavily methylated regions. MTases can be used to modify unmethylated DNA thereby enriching this fraction of the genome for analysis by sequencing.

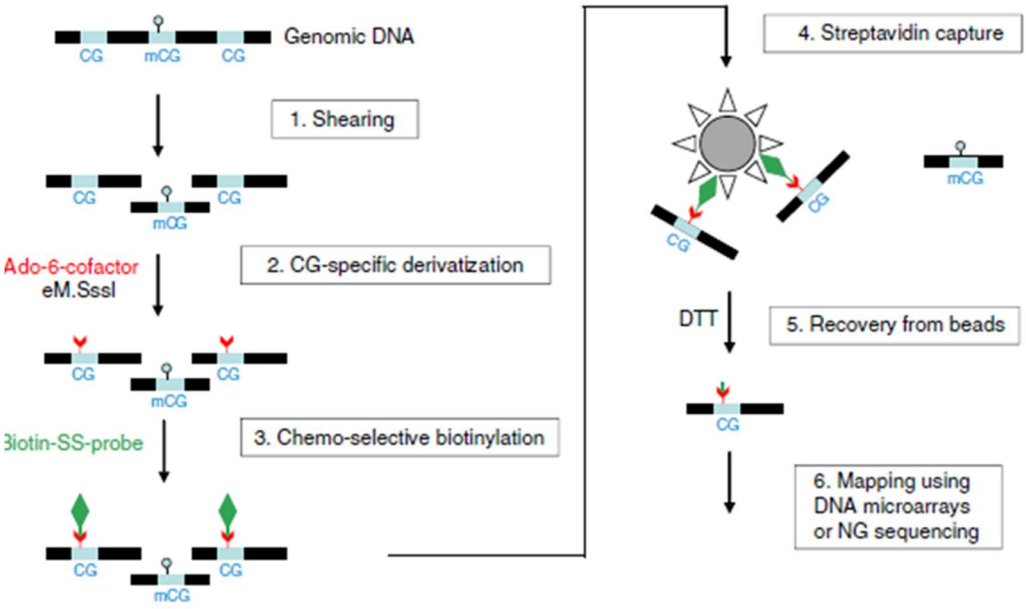


Fig. 14 mTAG labelling-based enrichment and analysis of unmethylated CpG sites in the genome. (a) Flow diagram of the analytical procedure. gDNA is randomly sheared to short fragments (Step 1) and treated with an engineered SssI DNA methyltransferase (eM.SssI) and a cofactor analogue to attach reactive groups to unmodified CpG sites (Step 2). The derivatized target sites are biotin-tagged using N-hydroxysuccinimidyl ester (Biotin-SS-NHS) (Step 3) and labelled fragments are selectively captured on streptavidin-coated magnetic beads (Step 4). Bound DNA fragments are recovered by cleavage of a disulphide bond in the biotin linker with DTT (Step 5). The enriched fragments are ligated to adaptors and PCR-amplified (Step 6) for microarray analysis or DNA sequencing (Step 7). (from E. Kriukiené et al., *Nature communications*, 2013.)

I.2.4. AdoMet dependent enzymes as therapeutic targets

Epigenetic modifications to DNA and proteins are important in regulation of many cellular processes. Abnormal levels of modifications, mutations to the epigenetic regulating proteins or dysregulation of their expression leads to a plethora of diseases including cancer. Hence, in recent years DNA methyltransferases as well as histone modifying enzymes, e.g. methyltransferases or deacetylases, have become important therapeutic targets.[150, 161] The methylation product – AdoHcy – acts as a non-selective feedback inhibitor for various MTases.[162] Hence, small molecule inhibitors are commonly designed to be an active inhibitor of MTase enzyme(s) or inhibit their activity indirectly by increasing the concentration of natural inhibitor – AdoHcy. This can be achieved by inhibition of *S*-Adenosylhomocysteine hydrolase enzyme which is responsible for the hydrolysis of AdoHcy in cells. A downfall of this approach is that AdoHcy is built up in cells and inhibits MTases non-selectively, potentially impacting many biological pathways in an uncontrollable way. Therefore, a great effort was put in the design and synthesis of selective inhibitors of AdoMet dependent enzymes.[150]. Although a range of non-nucleoside based small molecules were developed as potential inhibitors of AdoMet dependent enzymes, only nucleoside based and especially AdoHcy analogues will be presented.

There are a few nucleoside analogues that act as non-specific inhibitors of MTases with most known being Sinefungin and Methylthioadenosine (MTA) - Fig. 15. MTA is a natural nucleoside analogue of AdoHcy which similarly has pan-inhibitory properties. It is debated whether acts as a direct inhibitor of MTases or block activity of *S*-Adenosylhomocysteine hydrolase. Sinefungin is also a nucleoside analogue of

AdoHcy which is able to inhibit a range of MTases. Several derivatives of sinefungin have been introduced to target specific MTases.[163]

Fig. 15 Non-selective inhibitors of AdoMet dependent enzymes.

I.2.4.1. Inhibitors of protein methyltransferase enzymes

In 2012, the Luo group designed and synthesised a range of sinefungin analogues in order to target selectively protein methyltransferases. A n-propyl sinefungin derivative was identified as a promising inhibitor selective to SETD2 – a PKMT class enzyme - which aberrant activities are involved in developmental syndromes and cancers (Fig. 16 A).[164] Later, an analogue of sinefungin, bearing a cyclohexyl group in place of amino acid, was identified as an inhibitor of two PKMT enzymes – EHMT1 and EHMT2 (Fig. 16 B). These enzymes are responsible for mono- and demethylation of lysine 9 in histone 3 (H3K9) in euchromatin as well as other nonhistone substrates. Their malfunction leads to cancer, inflammatory diseases and neurogenerative disorders.[163]

AdoHcy analogues, due to the resemblance to AdoHcy structure, usually show low inhibition selectivity. Therefore, in order to improve selectivity an approach to design bisubstrate inhibitors can be made. This takes advantage of the two binding pockets in the majority of MTases – the AdoMet (cofactor) binding pocket, and that of the substrate (e.g. DNA, protein). The idea is to design a molecule that can block MTase activity by binding to both of the pockets simultaneously. The structural differences

between these regions in various MTases decreases the possibility of one molecule inhibiting a large group of MTases. This strategy led to development of many inhibitors including an inhibitor, bearing a guanidinium moiety, of G9a enzyme from the PKMT family, responsible for e.g. methylation of H3K9, which overexpression can cause cancer and neurological diseases (Fig. 16 C). However, its urea derivative showed inhibitory activity towards PRMT5 and PRMT7 enzymes which are a key players, working in cooperation, in monomethylation and symmetrical dimethylation of arginine residues in proteins (Fig. 16 D). In addition PRMT5 shows activity as a H4R3 methyltransferase which generates a binding site for DNMT3A on the nucleosome.[165] This could possibly be exploited to target DNMT3A activity.

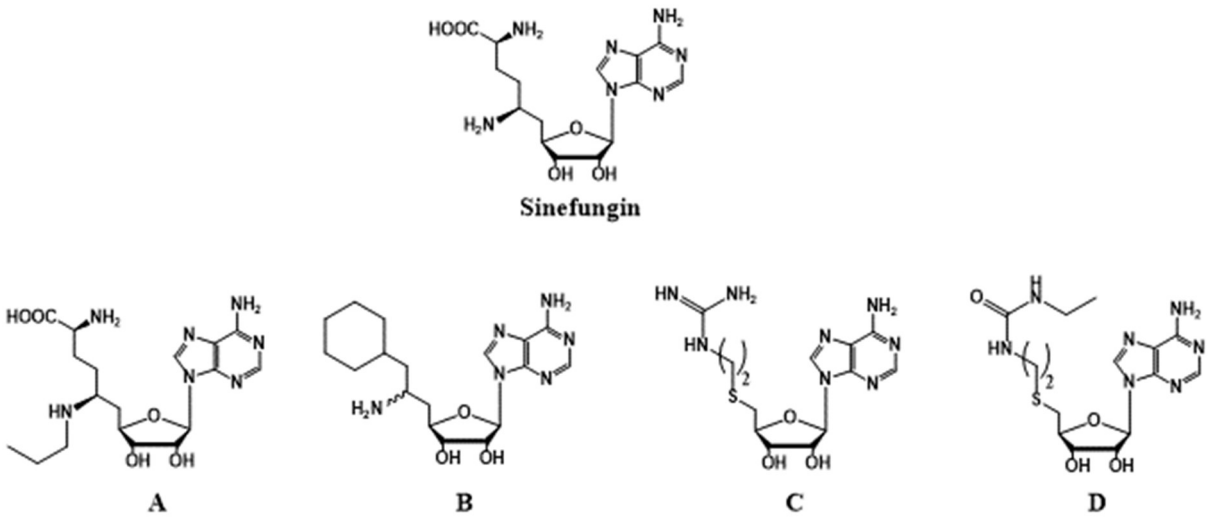


Fig. 16 Small molecule nucleoside based inhibitors of protein methyltransferases.

I.2.4.2. Inhibitors of DNA methyltransferase enzymes

DNA methyltransferases have been recognised as promising anti-cancer targets. Two nucleoside inhibitors, 5-azacytidine – Azacitidine or Vidaza; and its deoxy derivative 5-aza-2'-deoxycytidine – Decitabine, have been identified as DNMTs inhibitors and

approved for treatment of different types of blood cancer including myelodysplastic syndromes and acute myeloid leukemia (Fig. 17).

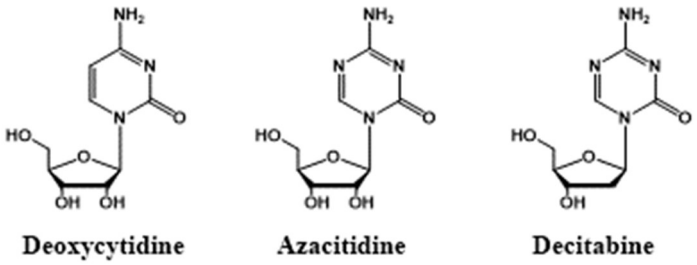


Fig. 17 Structures of natural cytidine and synthetic analogues azacytidine and decitabine

Both analogues are chemical analogues of deoxycytidine hence in cells they can be transformed into triphosphates and incorporated into DNA (Azacytidine also into RNA). As a result, both analogues can be recognised by DNMTs as cytosine and upon methylation, due to the substitution of carbon atom with nitrogen in 5-position, the β-elimination step is hindered. As a result DNMT is covalently bound to DNA which triggers a DNA damage response leading to degradation of DNMT (Fig. 18).[166] Unfortunately due to the low selectivity as well as high toxicity of both molecules there is a need for new DNMT inhibitors.

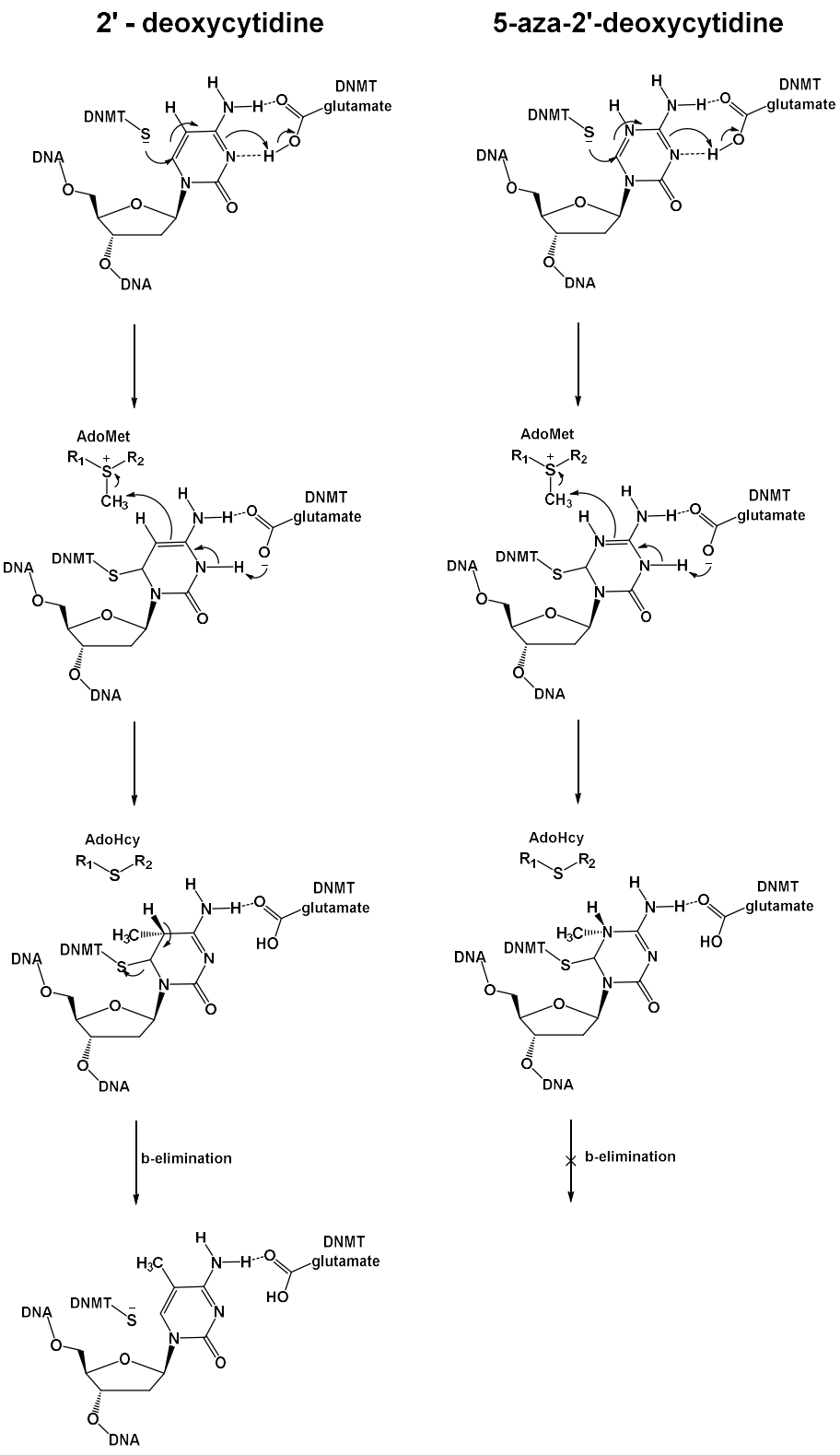


Fig. 18 Mechanism of DNMT inhibition by 5-aza-2'-deoxycytidine

Several positions on AdoHcy have been identified as important in AdoHcy inhibitory activity, hence they are interesting targets to introduce modifications. This includes an amino acid chain and adenine. Within adenine positions 2-, 7- and 6- are of most importance and well as substitution of entire adenine with other heterocycles. A few modifications to the 6-position of adenine were identified to maintain a moderate inhibitory activity but a high specificity towards DNMT3b over DNMT1 (up to 20-fold) – (Fig. 19).[167]

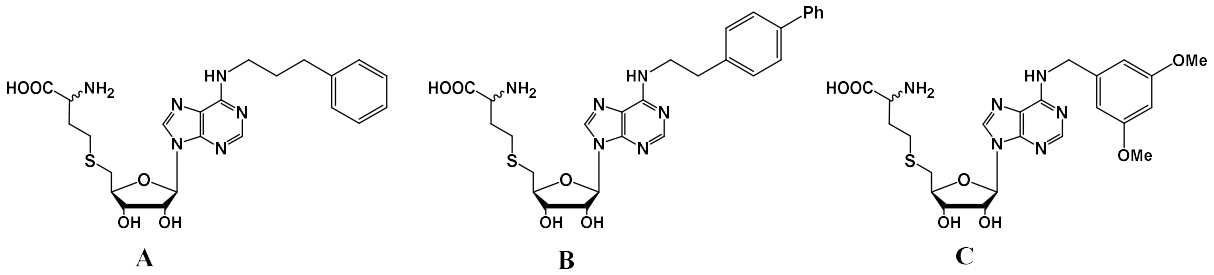


Fig. 19 N^6-modified AdoHcy analogues as DNMTs inhibitors.

I.2.4.3. Inhibitors of RNA methyltrasferase enzymes

In addition to DNA and protein methyltransferases, RNA methyltransferase enzymes are interesting target for antimicrobial treatment, particularly in antiviral treatment by inhibition of mRNA cap methyltransferases. Cap consists of a methylated guanosine residue located at the 5'-end which is important in the initiation of translation as well as protection of mRNA against ribonucleases. Such enzymes, usually unique for each virus, are found in Flaviviruses, e.g. Dengue virus, and coronaviruses e.g. SARS-CoV, MERS-CoV and are crucial in virus replication.[168, 169] A few N^6-substituted AdoHcy

analogues were identified as potent inhibitors against mRNA MTase in Dengue virus with increased specificity to RNA MTases compared to AdoHcy (Fig. 20).[168]

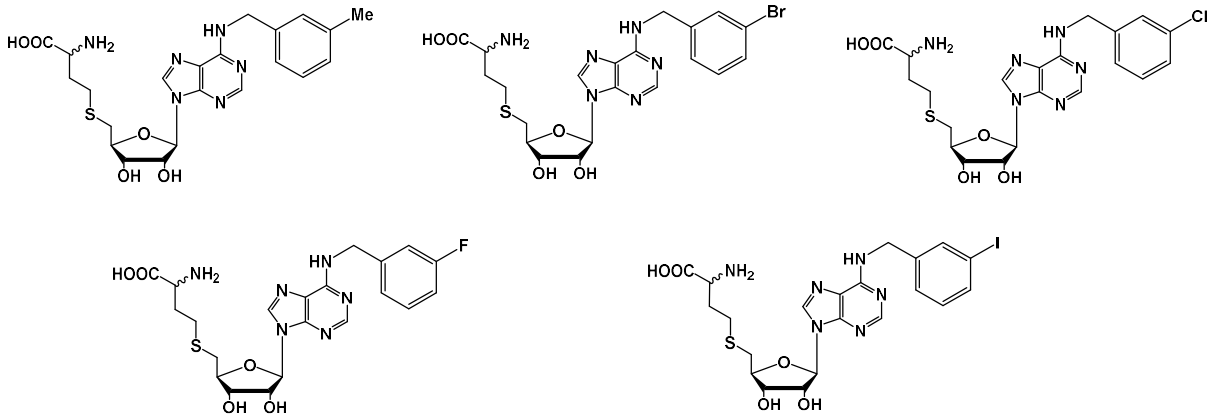


Fig. 20 AdoHcy analogues with specific inhibitory activity towards RNA methyltransferase enzymes.

I.3. Fluorescence microscopy

I.3.1. Fluorescence and fluorophores

Microscopy has been a dynamically developing field in science over the past several decades. It has found itself in numerous applications across many fields of science such as material physics, chemistry, molecular biology and medicine. It was used to study objects as small as single atoms, using scanning transmission electron microscopy,[170] to a self-organisation of biomolecules, cells or tissues using optical microscopy techniques. Most of the optical microscopy techniques use detection method of the reflected or transmitted visible light through the specimen. This usually results in an image of the entire sample with various structures present. It may be problematic in the selective studies of small objects where their signal can be difficult to distinguish from the multitude of present objects. A different principle underlies fluorescence microscopy which is based on the detection of emitted light from the object of interest.[171] This technique aims to record an image of fluorescent molecules ideally on the dark, non-fluorescent background. Theoretically, a single molecule can be resolved if there is a sufficient signal to noise ratio in the image. Unfortunately, it is not usually the case especially when working with biological samples such as cells or tissues which exhibit autofluorescence. This is a naturally occurring fluorescence, with emission wavelengths of ~350-550 nm, of many small molecule compounds bearing cyclic rings e.g. nicotinamide adenine dinucleotide phosphate ($NADP^+$) or riboflavin, proteins especially with aromatic amino acids e.g. collagen and elastin or organelles such as mitochondria.[172] Although autofluorescence has found some applications e.g. in the diagnostics of cancer where normal and transformed tissues show different patterns,[173, 174] in most cases it is an unwanted artifact when imaging specific

molecules or structures. Therefore, objects of interest are usually labelled with fluorescent molecules.

Fluorophores are usually organic compounds bearing a conjugated system of double bonds (Fig. 21A). In such systems, the energetic difference between the excited and ground electronic states is small enough that the electrons can be transferred onto higher energy level upon absorption of UV or visible light. The energy of the excited electron can be dissipates in a variety of different ways. Kinetically, via vibrational relaxation or, in some molecules, with no energy lost. Electron transfer is another non-radiative relaxation mechanism, with the change of spin switching to the forbidden triplet state. Another way is via the return of an electron onto a ground state with the emission of, a typically lower energy than that absorbed, photon (Fig. 21B). If the electron was in a singlet state the process is fast (10^{-9} s) and called fluorescence otherwise a transmission from triplet state is called phosphorescence and usually is a thousand times slower (10^{-6} s) which is important in some super-resolution imaging techniques (See I.3.5).[171, 175] A small molecule fluorophore can be covalently attached to the functionalized biomolecules, i.e. proteins or DNA, via many chemistries including NHS or click reactions.[176, 177]

Some of these synthetic fluorophores do not require conjugation to a biomolecule or a structure of interest and have shown the ability to accumulate in specific organelles such as mitochondria, nuclei or lysosomes.[178-180] In addition, some fluorophores with environmentally sensitive absorption or emission properties can be used as sensors of ion concentration, e.g. calcium or hydrogen, or transmembrane voltage in the studies of neuronal activity.[181-183] Organic fluorophores are widely used as conjugates with

antibodies in immunofluorescence technique. Such a conjugate is used to stain proteins or other biomolecules recognized by the antibody.

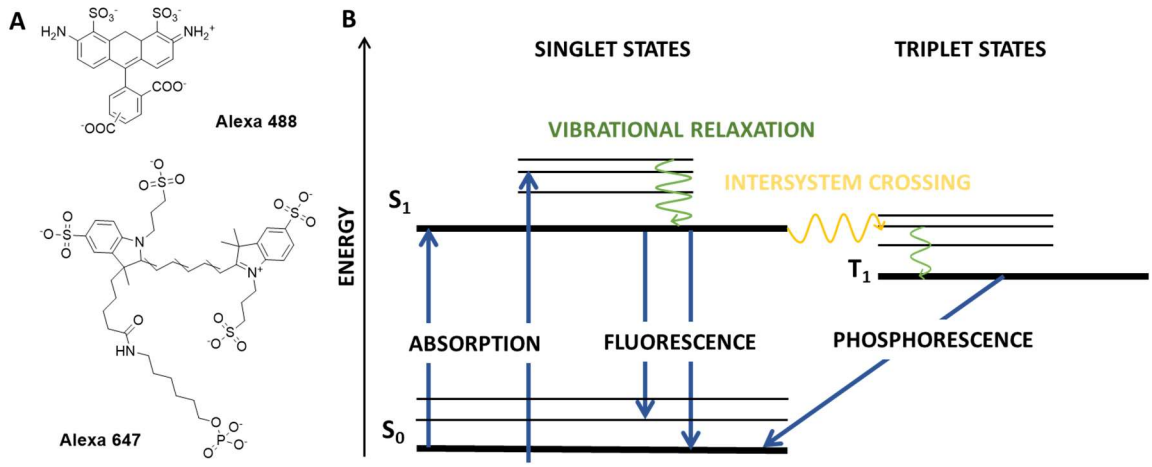


Fig. 21 Structures of common fluorophores Alexa 488 and Alexa 647 (**A**) and a simplified Jablonski diagram showing the energy states and transitions in a molecule. In general molecules occupy the ground state S_0. Upon the absorption of photons, electrons can excited to the higher energy state (S_1) or its vibrational states of S_1 (thinner horizontal lines). Then the energy can be dissipated in various ways such as kinetic in vibrational relaxation (green curved lines) or with spin change entering triplet state via intersystem crossing (yellow curved line). Electron can also emit a photon either from singlet state (fluorescence) or triplet state (phosphorescence).

In addition to synthetic fluorophores, some natural fluorescent systems have been identified. A protein extracted and identified in 1962 by Shimomura *et al*. from the jelly fish *Aequorea victoria* known as GFP (Green Fluorescent Protein) contains a conjugated system of double bonds and emits light in green region (max. 509 nm).[184] Since then, many mutant proteins of GFP and others have been developed to cover a broad range of emission wavelengths (from blue to red).[185, 186] Fluorescent proteins can be encoded in such a way that are co-expressed in cells as a hybrid with a cellular target. In the ideal case, fluorescent labelling allows the concentration, distribution, expression levels and structure of the target molecules to be studied in living cells.[187]

I.3.2. Widefield fluorescence microscopy

In fluorescence microscopy, there are a range of microscopes which can be adjusted to various applications. The most common setup, due to its simplicity, is the epi-fluorescence (widefield) microscope (Fig. 22). In this configuration a single objective is used to magnify, focus the excitation beam onto the specimen and also collect the emitted signal from the sample. The light from the source, e.g. a laser, is directed to the objective on dichroic mirror. After illumination, emitted light from fluorophores and some reflected excitatory light are collected by the objective. On their way back the light is transmitted through the dichroic mirror and emitted wavelength of interest is filtered out before reaching the detector.[171]

The widefield microscope has many advantages such as a lower cost of the microscope compared to more sophisticated setups, a wide field of view which can be registered at once, and a fast acquisition of the images which especially enables three-dimensional objects to be imaged within seconds and directly observed. Therefore, it is usually the first choice fluorescence microscopy method for life sciences. Unfortunately, the simplicity comes at the cost of image quality limitations. Due to the fact that a wide section of specimen is excited simultaneously, a high level of signal originating from out of the focal plane fluorescence is detected.[188] This results in blurring of the image, low contrast and can limit spatial resolution. This effect on the resolution adds to the limit of the diffraction of light which is about 200-300 nm in the lateral direction and 500-700 nm in the Z-axis.[189] Therefore, objects closer than the aforementioned limits cannot be resolved using widefield microscope.

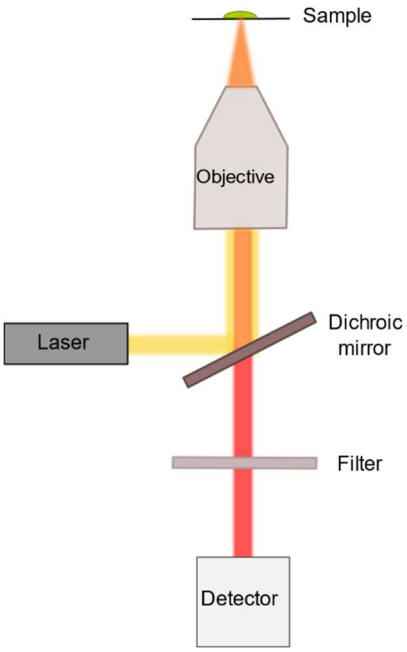


Fig. 22 The inverted epi-illumination fluorescence microscope.

I.3.3. Deconvolution algorithm

One of the methods to improve the resolution and blurring issues is the deconvolution algorithm. Deconvolution methods aim to predict the out-of-focus signal within one section of the image and remove it.[190] In order to identify the contribution of a background signal (blurring, convolution) one needs to characterize the point spread

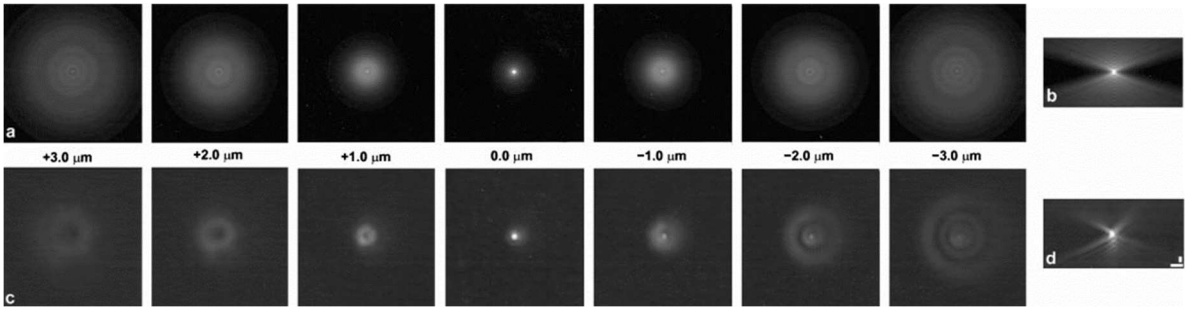


Fig. 23 Sequential focal planes through a theoretically predicted (a) or experimentally determined (b) point-spread function (PSF) for a 100x, 1.35 NA Olympus UplanApo objective. Distances above and below focus are shown from +3.0 mm to -3.0 μm. Observe that for the theoretical PSF (a) rings increase in number and grow in diameter as the point source is imaged above or below focus (0.0 mm corresponds to in focus). Ring patterns in (a) are symmetrical at equal distances above and below focus, whereas for the experimental PSF (c), ring patterns are more pronounced below focus. When viewed from the side (an XZ view), the PSF forms a double cone (b, theoretical PSF; d, experimental PSF). The edge of the cone corresponds to the increasing ring size as the microscope focal plane moves further away from the point source. The presence or absence of symmetry in ring patterns is also evident in the XZ view. For the experimental PSF (d), this asymmetry in ring patterns arises from spherical aberration, most likely a defect in this particular objective. Intensities in a–i are displayed on a logarithmic scale to highlight the weak outer rings of the PSF. Bars: 1 μm. (from Mcnally et al. *Methods*, 1999)

function (PSF) of the imaging system. The PSF is the theoretical image of a single point object, produced by an imaging system and is characteristic to a specific microscope setup. The PSF results from the processes of diffraction or aberration in the lens of the system, whereby the image of a perfect single point object to be blurred and occupy a larger volume[191] The PSF can be predicted theoretically and modelled using available programmes, such as plugins in a common open-source software *Fiji*,[192] knowing parameters of the imaging experiment such as refractive index of the medium, numerical aperture of the objective and wavelength of the emitted light. The PSF can also be measured by taking Z-stack image of the point objects close to or below the resolution limit e.g. fluorescent spheres.[193] An example of predicted and experimental PSFs is shown (Fig. 23). In this case, the image of a fluorescent particle in focus (0 μm) is circular in both cases. When moving further and further away from the focus plane the signal shows a pattern of bright and dark circles (Fig. 23 A, B). Although some asymmetry can be noticed in the experimental PSF (Fig. 23 D) which can be improved by adjusting the setup, the theoretical PSF gives quite a good representation. In the deconvolution process, once the PSF is identified, the 3D image of the specimen is taken and any inhomogeneous illumination such as random single pixel spikes are corrected. In the next step, the mathematical algorithm is applied to the recorded image. An acquired light distribution (an image) *I(X,Y,Z)* of the specimen can be represented as a combination of a three dimension light distribution of the specimen *S(X,Y,Z)* and the known distribution of *PSF(X,Y,Z)* which can be mathematically described as:

$$I(X,Y,Z) = S(X,Y,Z) \otimes PSF(X,Y,Z)$$

Knowing this relation, in each point of the image, sums of the contributions of PSFs from other volumes can be determined and the above equation can be solved for *S*. There are many methods available to solve the above equation which in practice might be a computationally demanding process.[193]

As a result, a deconvolved image usually has improved contrast and reduced background signal with the edges of features sharper (Fig. 24).[192] Although the deconvolution is a useful operation to overcome the issues of simple widefield fluorescence microscopy, care during the interpretation of the images must be taken. Even if the PSF is characterized accurately, the deconvolution may introduce artifacts

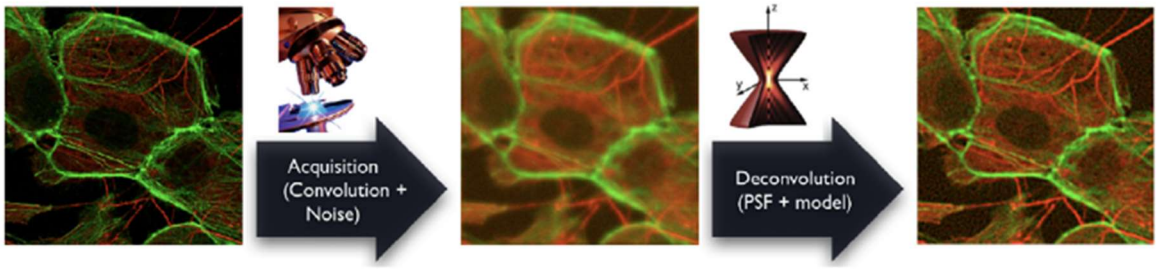


Fig. 24 Principle of the deconvolution of a z-stack of images. (from D. Sage et al. *Methods*, 2017)

or delete some information. Therefore, when analysing the deconvolved data one should always be suspicious when a structure appears in deconvolved image with no traces of it being present in the pre-processed image. Moreover, as there is not an optimal (one-size-fits-all) deconvolution method, several approaches should be tested on the same set of data to identify the method specific artifacts. The usual artifacts that can be found in such processed images are periodic patterns e.g. circles or rings usually around bright objects or elongation of the objects in the Z-direction.[193] Even though there are some limitations and drawbacks of the deconvolution method it is a method that can be applied to the images already taken using a simple and fast

widefield microscope. However, the resolution of the image is still limited which is more pronounced in imaging of three-dimensional structures.

I.3.4. Selective plane illumination microscopy

Selective plane illumination microscopy (SPIM) was introduced in 2004 by Huisken et al. to image live fish embryos.[194] This technique overcomes the issue of high out-of-focus plane signal by decoupling illumination and detection channels. In principle, the illumination and detection lenses are located at 90° angle to each other compared to the widefield microscopy where the same lens is used for excitation and detection (Fig. 25 A). In addition, the excitation light path is modified by a cylindrical lens in such a way that forms a sheet of light illuminating only a thin plane of the sample. Therefore, the other objective detects only the light from the illuminated plane which significantly reduces the out-of-focus plane fluorescence.

In addition to the lowered background fluorescence, SPIM also greatly decreases photodamage of the sample compared to the epi-fluorescence method. In widefield microscopy even though the in-focus object is imaged once at a time, the entire volume of the sample above and below is illuminated. This can cause a photobleaching of these regions, decreasing the quality of the volumetric image, but also in the case of the imaging of living biological samples, prolonged exposure can cause a photodamage. In SPIM the volume prone to photodamage is greatly decreased due to the illumination of the small volumes that are imaged at a time (Fig. 25 B). Therefore, SPIM has been used to image living organisms over long periods of time (several hours) without inducing a significant amount of photodamage.[194-196]

Similarly to widefield microscopy, SPIM can be used to quickly produce three-dimensional images. Stacks of images in parallel planes can be taken and easily reconstructed to give a three-dimensional representation of the imaged structure. When imaging the biological samples, the light in the excitation path can be absorbed or scattered causing a decrease in the quality of the images. This is especially noticeable in structures located further away from the illumination lens. Therefore, in the initial setup, the objectives were located horizontally and sample was located in the rotatable agarose cylinder in which the biological sample in an aqueous media was placed. This allowed the image of the sample to be taken from multiple views and improve the quality of the image throughout the sample, but the setup was complicated and not compatible with standard sample preparation e.g. on the glass slide. Therefore, modifications have been introduced and objective were positioned vertically so the sample can be imaged from the top. This allowed the SPIM extension to be compatible with commercially available inverted microscopes in which the objective in widefield microscopy is located below the sample. The method was then called *inverted* SPIM (iSPIM) (Fig. 25 A).[197]

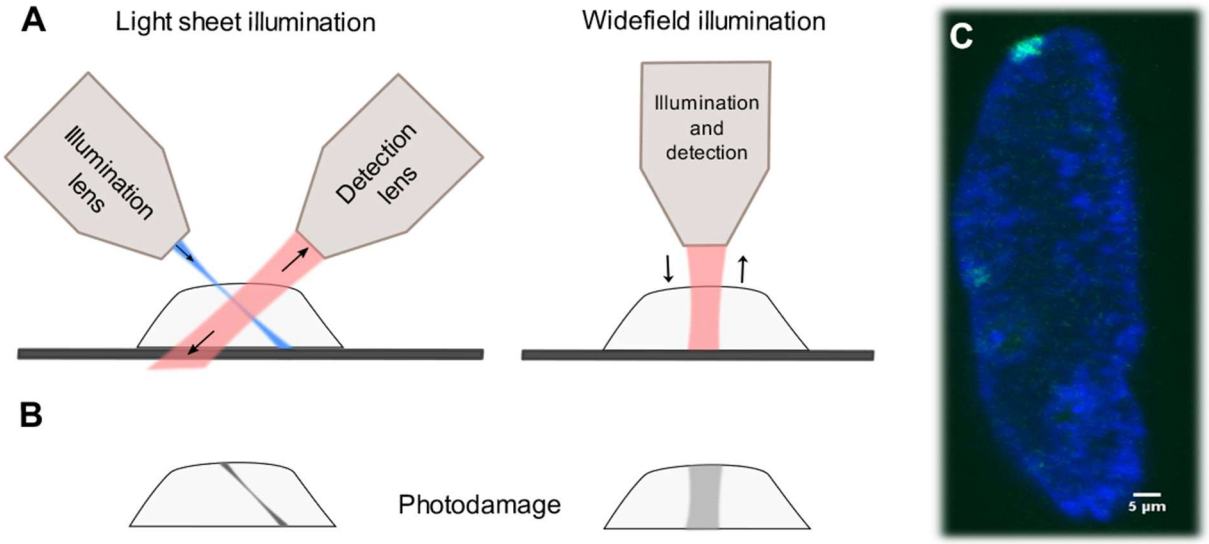


Fig. 25 Schematic of the inverted selective plane illumination microscopy (iSPIM) against widefield microscopy (**A**). The comparison of photodamage induced to the sample by SPIM and widefield microscopy (**B**). A single plane image of MCF10A cells in the expansion microscopy gel acquired on SPIM (**C**).

Although the iSPIM alone improves the signal to noise ratio and decreases photodamage of the sample whilst keeping the high speeds of volumetric imaging, there are some limitations to the method. Due to the instrument setup, the flexibility of the sample preparation is limited. The non-intuitive geometry makes the navigation within the sample quite challenging. Another difficulty that one may encounter using SPIM, is the large size of the data generated which may be as high as 1 GB s^{-1}.[198] This complicates the data storage, transfer and processing. In standard iSPIM method the resolution of the images is similar to conventional microscopes and is limited by the diffraction of light. This can be overcome by coupling light sheet microscopy with super resolution microscopy methods.[199-201]

I.3.5. Beyond the diffraction limit – briefly about super resolution microscopy

Many microscope techniques aiming to improve the contrast and resolution of the images, using numerous approaches, have been developed over last decades. Confocal microscopy applies the detection of light through the designed small pinhole

to eliminate out-of-focus light as much as possible (Fig. 26). Although the contrast of the image is improved the resolution is still limited to the size of the PSF.[202]

Another approach using different periodic illumination patterns which blend with structural patterns of the sample and shift their frequencies was applied in structured-illumination microscopy (SIM).[203, 204] The density and orientation of illumination patterns is varied producing different interference patterns with the sample in obtained images. From a set of multiple images, the applied known pattern can be extracted uncovering structures of the sample. SIM achieves a subdiffraction resolution both laterally (100 nm) and axially (300 nm)[204-206]

In order to improve the resolution even further PSF of the microscope can be manipulated. As an example, stimulated emission depletion microscope (STED) technique which in addition to the basic excitation beam applies a ring-shaped beam which stimulates emission of the excited fluorophores located in this outer area at lower energies. This effectively removes them from the recorded signal leaving the improved and sharper effective PSF in the centre (Fig. 26). This led to the greatly improved lateral resolution of ~60 nm.

The last presented approach is super-resolution fluorescence microscopy using single molecule localization. Over years there have been many techniques developed that are based on similar principle of the image reconstruction from many single point images. When a multiple single fluorophores are located close to each other and their fluorescence is recorded at the same time it is very difficult to resolve their location. Whereas when fluorophores, which fluorescence can be switched ‘on’ and ‘off’ are applied the image containing a greatly reduced number of active fluorophores can be

obtained. The information about single fluorophores localization taken from hundreds or thousands of images can be then combine to reconstruct a super-resolution image of the labelled structure (Fig. 26). There are various methods used to achieve the 'on' and 'off' fluorescence states. One of them is the properties of some fluorophores to enter the dark state i.e. the transition of the excited electron to a triplet state. This leads to even a thousand or longer times between excitation and emission of the light than in fluorescence process.[207] This concept was used in various techniques such as STORM, PALM or FPALM.[208-210] An effective lateral resolution of ~20 nm has been

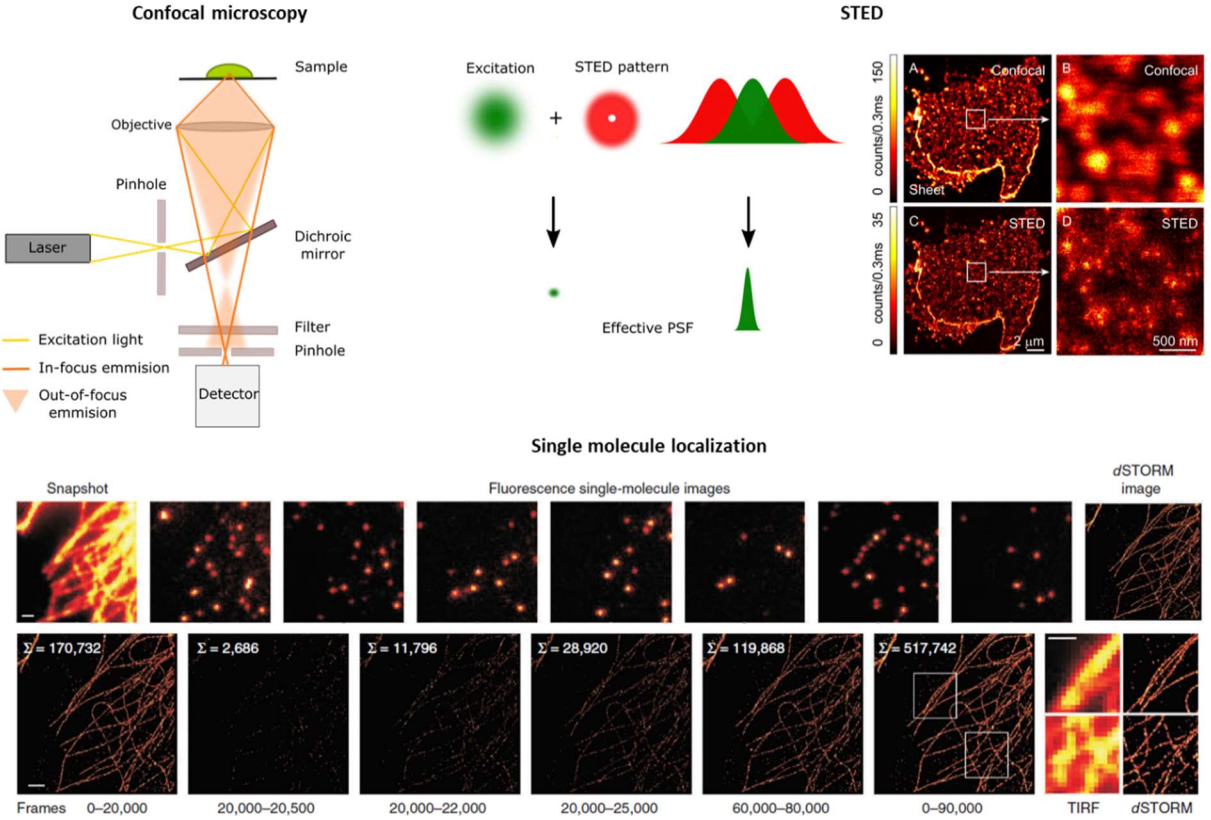


Fig. 26 Schemes showing principles of confocal microscopy, STED and single molecule localization based microscopy techniques. **STED** right panel: Acetylcholine receptor (AChR) spots and AChR nanoclusters in single plasma membrane sheets. Plasma membrane sheets were obtained by an ultrasound pulse applied to CHO-K1/A5 cells adhered to the glass coverslip, fixed, labeled and then imaged in the confocal (A, B) and STED (C, D) modes. Left column: the entire glass-adhered plasma membrane. Right column: high magnification detail of a 1.95×1.95 µm2 area illustrating the resolution of single confocal spots into multiple AChR nanoclusters with STED microscopy. From Kellner R. R., et al., Neuroscience, 2007. **Single molecule localization**: dSTORM acquisition procedure exemplified for Alexa Fluor 647–labeled β-tubulin in a COS-7 cell. (**top**) A fluorescence image of the structure is measured at low excitation intensity. (**bottom**) Image reconstruction showing that a highly resolved image can already be reconstructed from 20,000 images corresponding to a total acquisition time of ~20 s. The structure cannot be fully resolved after analysis of <2,000 frames. Σ, the number of localizations used to reconstruct the dSTORM image. Scale bars, 1 µm. Adapted from Van de Linde S., et al. Nature protocols, 2011

achieved with many variants of the technique used to visualize DNA-protein complexes and other cellular structures.

Although all of the presented techniques found many applications and all have sub-diffraction-limit resolution, they also have some limitations. Due to the limited detected light in some techniques (confocal, STED) a high intensity excitation light is needed which is usually not compatible with live imaging due to toxicity but also can lead to fast photobleaching of the fluorophores, even in fixed samples.[211, 212] A number of fluorophores compatible with these methods is also limited due to the varying properties required by these techniques.[213, 214] As some techniques, such as STED or single fluorophore localization based techniques, require large numbers of images to produce an image of a single structure, the visualization of large three-dimensional volumes is challenging and time consuming. In addition, a high cost and specialization of the instruments, hence restricted access may be an issue in frequent use.

II. RESULTS

II.1. Synthesis of N6-substituted AdoHcy/AdoMet analogues

II.1.1. Introduction

II.1.1.1. Applications of N6-substituted AdoHcy/AdoMet analogues

Derivatives of AdoHcy are potential inhibitors of AdoMet-dependent enzymes important in the development of various diseases including cancer or neurological diseases or as antiviral drugs (See I.2.4). In addition to pharmaceutical applications, modified AdoMet/AdoHcy analogues can serve as specific cofactors/inhibitors in the studies on activity of AdoMet-dependent enzymes.

Several atoms of the AdoMet/AdoHcy structure have been identified as promising for alteration of its properties, N6 position of adenine a common target for modification.[167, 215] Modification to the 6 position of adenine with the application of 'bump-and-hole' strategy, to accommodate a bulky modification in the binding pocket of an enzyme, led to the identification of N^6-benzyl-AdoMet analogue as a selective cofactor of mutant Rmt1 enzyme which is a part of PRMT family (Fig. 27). In addition, N^6-napthylmethyl-AdoHcy and N^6-benzyl-AdoHcy derivatives were reported as allele-specific inhibitors with >20 fold selectivity over wild-type enzyme.[158] Further studies of gene and mRNA expressions after treatment of yeast cells with these inhibitors showed the downregulation of both Rmt1 methyltransferase and CMK2 protein kinase, which suggest possible link between protein methylation and phosphorylation.

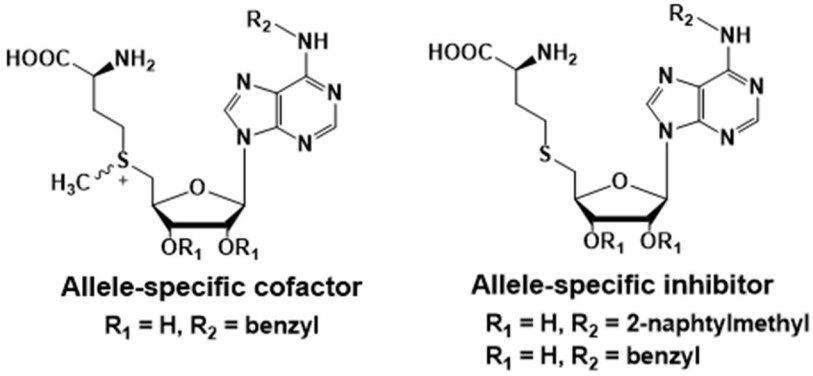


Fig. 27 AdoMet/AdoHcy analogues as allele-specific cofactors/inhibitors of Rmt1 enzyme.

The Lou group applied the 'bump-and-hole' strategy to find selective AdoMet analogue/mutant protein pairs e.g. Pob-SAM/PRMT1 Y39FM48G.[216] Pob-SAM (or here as AdoHcy-6-yne) analogue (Fig. 28) is not recognised by native PRMT1 which allowed them to later label PRMT1 substrate *in vivo* in presence of other MTases especially native PRMT1.[217]

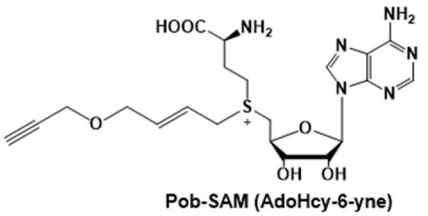


Fig. 28 Structure of Pob-SAM (AdoHcy-6-yne) specific cofactor to PRMT1 mutant

Modification of the N6 position of adenine can be used to alter the recognition of AdoMet/AdoHcy by a target MTase and also to introduce functionality into biomolecules such as proteins. This was exploited with aziridinoadenosines where the entire cofactor, bearing functionality such as propargyl- or azide groups in the adenine moiety, is transferred onto the substrate and can be further conjugated via click chemistry to e.g. biotin and function as epigenetic probes of such biomolecules.[218]

Moreover, AdoHcy/AdoMet analogues can also be used as small molecule probes in AdoMet-binding proteins studies. Complex, N6-modified AdoHcy analogues bearing a capture motif i.e. biotin, crosslinking group e.g. azide were synthesized by the Weinhold group (Fig. 29).[215] These analogues were used as capture compounds for MS studies of AdoHcy/AdoMet binding proteins and allowed the capture of AdoMet/AdoHcy-binding proteins from *E.* Coli lysate.

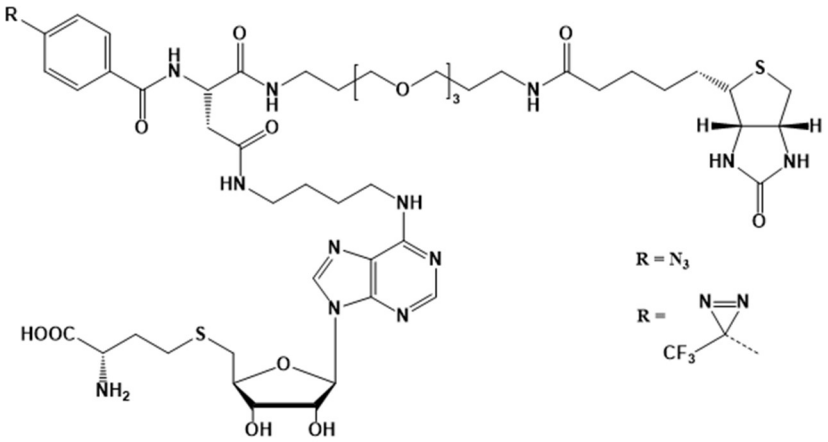


Fig. 29 N^6-substituted AdoHcy analogue for capturing of AdoHcy/AdoMet binding proteins

II.1.1.2. Current strategies of synthesis of N^6-substituted AdoHcy/AdoMet analogues

N^6-substituted AdoHcy/AdoMet analogues are versatile tools for inhibition of AdoHcy/AdoMet dependent enzymes, specific cofactors or capture compounds. However, there are limited synthetic strategies to introduce modifications to the 6-position of adenine (Fig. 30). Usually 6-chloropurine riboside (6-Cl-Ade) is used as chlorine atoms can be easily replaced via a nucleophilic aromatic substitution (S_NAr) by groups of varied in polarity and size (Fig. 30 A and B a)). Next step involves

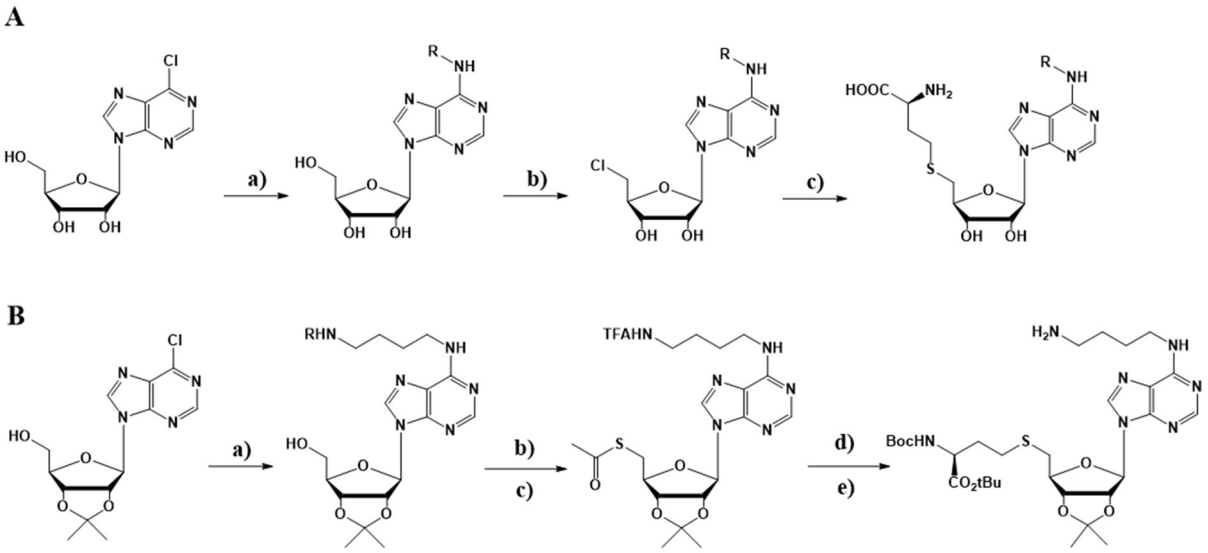


Fig. 30 Synthesis strategies of N^6-substituted AdoHcy analogues. A – a) RNH_2, EtOH; b) $SOCl_2$/HMPA; c) L-homocysteine sodium salt, KI, H_2O, reflux; B – a) 1,4-diaminobutane, NEt_3, EtOH, 60°C, 96%; b) ethyl trifluoroacetate, NEt_3, MeOH, room temperature, 92 %; c) PPh3, DEAD, AcSH, THF, 10°C then 0°C, 94%; d) 1: NH_3, MeOH, 0°C; 2: N-Boc-γ -tosyl-homoserine-tert-butylester, KOH, 18-crown-6, THF, 10°C, 79% for both steps; e) NH_3, H_2O, MeOH, room temperature, 90%

introduction of L-homocysteine at the 5'-position of the ribose ring. This can be done via two routes. The first requires activation of 5'-OH group by selective chlorination e.g. with thionyl chloride followed by a nucleophilic substitution with L-homocysteine (Fig. 30 A b and c).[158] The second option is introduction of sulfur with thioacetic acid under Mitsunobu conditions (Fig. 30 B). The resulting thioester is then cleaved and reacted with tosyl derivative of protected homoserine. The first route, does not allow modifications that could react during the chlorination step such as primary -OH groups. Hence, it is usually used to introduce non-polar inert groups such as methyl, benzyl etc. On the other hand, the second route allows introduction of functional modifications bearing terminal groups, such as amine. However, due to the lack of -OH selectivity during the thioester formation, this method requires protection of 2'- 3'-OH groups on the ribose ring. The synthesis of diamine modified AdoHcy was achieved in 6 steps,

required several protecting groups and resulted in semi-deprotected compounds with a total yield of 59 %.[215]

Therefore, we decided to design a fast and efficient synthesis method of N^6-substituted-AdoHcy/AdoMet analogues that does not require protecting groups and allows the introduction of a wide range of N6-modifications.

II.1.2. Results and discussion

II.1.2.1. Optimisation of the synthesis of N^6-substituted AdoHcy/AdoMet analogues

In order to tackle current limitations of the synthesis of N^6-substituted AdoHcy analogues we decided to follow the synthetic path A with chlorination step as it does not require protecting groups on ribose (Fig. 30). However, the chlorination step using thionyl chloride requires a protection of some chemical groups, e.g. -OH or -COOH, introduced in the 6 position of adenine in the previous step. Therefore, we decided to perform activation of 5'-OH group prior to the introduction of the modification in the adenine ring. As a result we would obtain a doubly-activated intermediate 5',6-

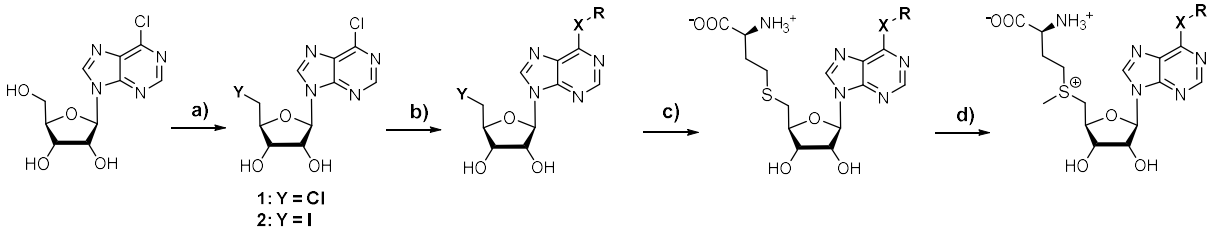


Fig. 31 Synthetic scheme of N^6-substituted AdoHcy/AdoMet analogues. Reaction conditions: (a) for 1: $SOCl_2$, pyridine, acetonitrile, 24h, 80%; 2: I_2 , PPh_3, imidazole, NMP, 24h, 76%; (b) linker, NEt_3, water/MeOH, 46-95%; (c) L-homocysteine, 1M NaOH, MeOH 100°C, 35-98%; (d) CH_3I, $AgClO_4$, $HCOOH/CH_3COOH$ (1:1), 30°C, 35-69%;

dichloropurine 5'-deoxyriboside (5',6-diCl-Ade) - Fig. 31. Due to the difference in reactivity of 5'-Cl and 6-Cl in nucleophilic substitution reaction, substitution of 6-Cl can be carried out at room temperature whereas 5'-Cl requires high temperature (100°C), we hoped to be able to selectively introduce a functionality in the nucleobase (Fig. 31

b). An intermediate that can be selectively modified is ideal for a synthesis of a library of modified AdoHcy/AdoMet analogues as 5'-OH activation step could be performed at the beginning of the synthetic process and subsequently used to introduce a range of moieties.

The second crucial step during the synthesis of AdoHcy/AdoMet analogues is coupling with L-homocysteine (Fig. 31 C). Low reactivity of 5'-Cl requires high temperature (100°C) and, typically, long reaction times (~24 hours) under basic conditions. Such conditions could possibly be damaging to temperature sensitive modifications. Hence, we decided to introduce a more reactive iodine in place of the 5'-Cl substituent which should improve the rate of the coupling reaction. However, as alkyl iodides are about two orders of magnitude more reactive compared to chlorides,[219] there was a danger that this approach may decrease the selectivity of introduction of the modifications to the nucleobase over 5' position in the ribose.

II.1.2.2. Synthesis of 5',6-diCl-Ade and 5'-I,6-Cl-Ade

The first step of the designed synthesis path of N^6-substituted AdoHcy/AdoMet analogues was the preparation of intermediate which is either 5'-Cl or 5'-I activated 6-chloropurine riboside (Fig. 32).

Chlorination of the 5'-OH was obtained by selective activation of the 5'-OH through its nucleophilic substitution with $SOCl_2$ in acetonitrile in the presence of pyridine (Fig. 32 top).[220]

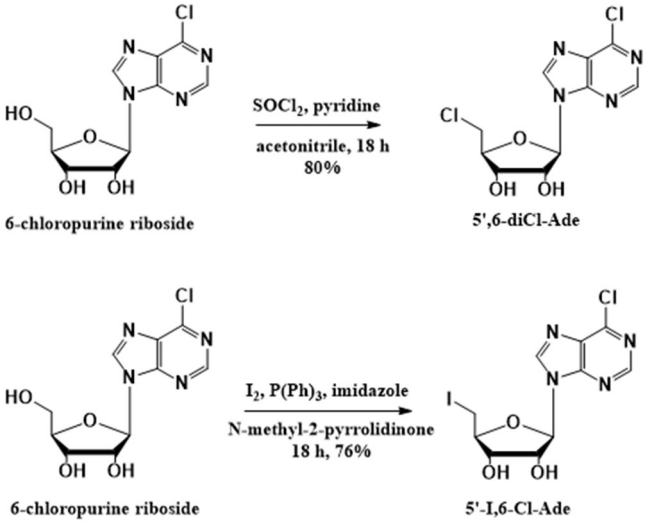


Fig. 32 Synthesis of 5’,6-diCl-Ade (top) and 5’-I,6-Cl-Ade (bottom).

The second intermediate 5’-iodo-6-chloropurine 5’-deoxyriboside (5’-I,6-Cl-Ade) was attempted by 5’-iodination method (Fig. 32 bottom) commonly used for the iodination of guanosine due to its convenient purification process. The activated guanosine forms a precipitate when poured into a DCM/water mixture and can be easily collected.[221]

This approach was implemented with 6-Cl-Ade as a starting material. However, pouring the reaction mixture into DCM/water resulted in no precipitate and the product was present in both layers despite keeping the mixture overnight in the fridge. This suggests that the oxygen atom at 6-position and amino group at 2 position in guanosine are crucial for precipitate to be created in the interlayer of the solvents mixture. In order to address this, crystallization of Ph_3PO from the crude reaction mixture, followed by purification of the product by preparative RP-HPLC was applied.

After successful synthesis of both of the intermediates with a good yields (80% and 76%), the selective introduction of functional modifications in the 6-position of the adenine was examined.

II.1.2.3. The introduction of functionality to the nucleobase

The next step in the synthesis of N^6-modified AdoHcy analogues in our approach is the attachment of various modifications to the nucleobase via nucleophilic substitution on the aromatic ring. We found that 6-Cl can be selectively substituted with an amine over 5'-Cl and 5'-I due to the mild conditions of this reaction (Fig. 33).

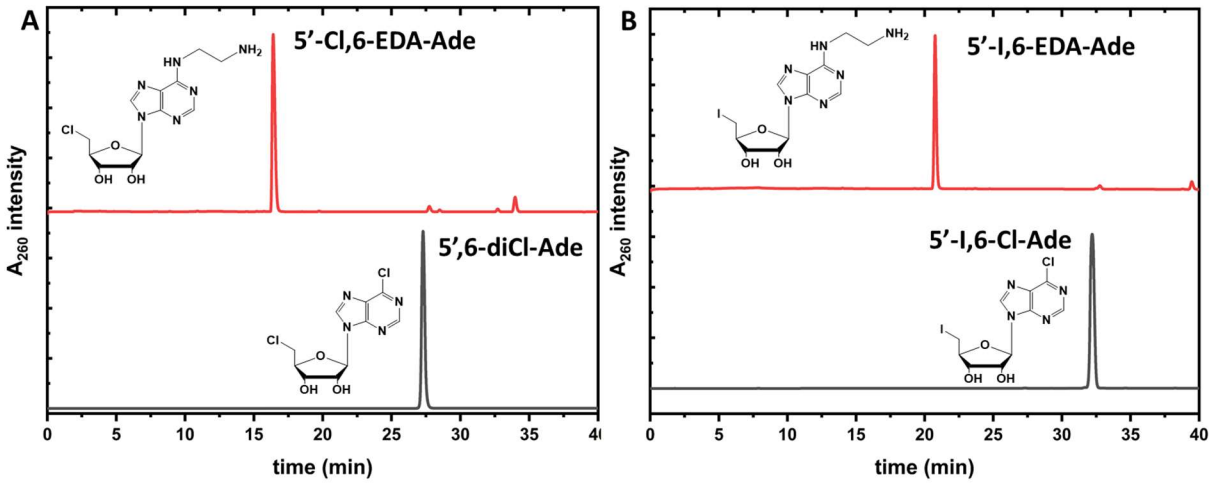


Fig. 33 HPLC profiles of ethylenediamine substitution to the 6 position of adenine in 5',6-diCl-Ade (**A**) and 5'-I,6-Cl-Ade (**B**) intermediates.

As a result, a wide range of amine based linkers bearing various functional terminal groups such as amine, carboxyl, azide, alkyne, alkene and hydroxyl were reacted with either 5',6-diCl-Ade or 5'-I,6-Cl-Ade intermediate (Fig. 34).

This reaction is efficient (>70 % yield) and is completed in 2-8 hours for most of the linkers. However, propargyl amine requires longer reaction time (~48 hours) due to lower nucleophilicity of the amine located close to the alkyne whereas a lower yield (46 %) was observed in the reaction with the secondary amine - L-proline.

This library of functionalised analogues of adenosine was subjected to the coupling with L-homocysteine. The reactivity of 5'-Cl and 5'-I activated analogues was compared.

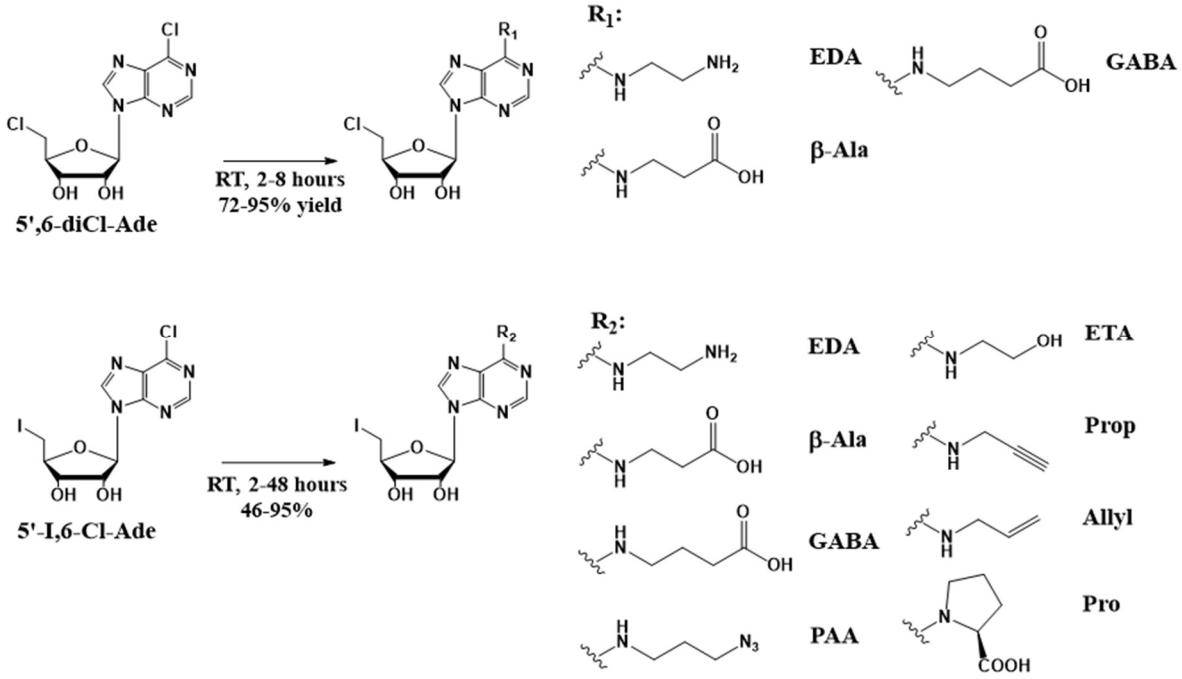


Fig. 34 Reaction schemes of introduction of functionalisations to the 6 position of adenine.

II.1.2.4. Comparison of the reactivity of 5'-Cl versus 5'-I activated functionalised adenosine derivatives with homocysteine

The routine synthesis of AdoHcy is through the 5'-Cl substitution with homocysteine in the presence of base. The coupling of activated nucleoside with homocysteine is usually the limiting step in the synthesis of AdoHcy/AdoMet analogues due to the high temperature required (100°C) and long reaction time (~24 hours). Therefore, we have reacted 5'-Cl activated analogues and their 5'-I counterparts with homocysteine to

determine whether the introduction of iodine to the ribose improves the speed of this reaction (Fig. 35).

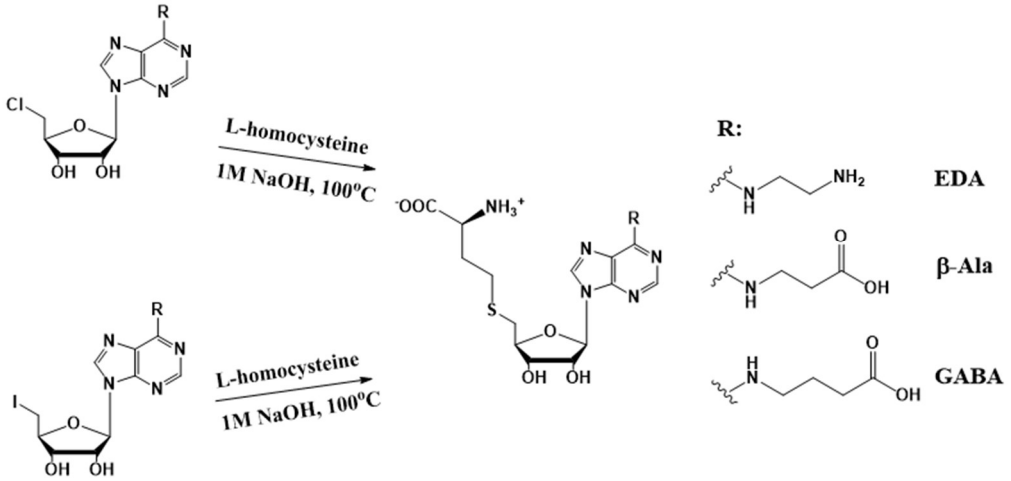
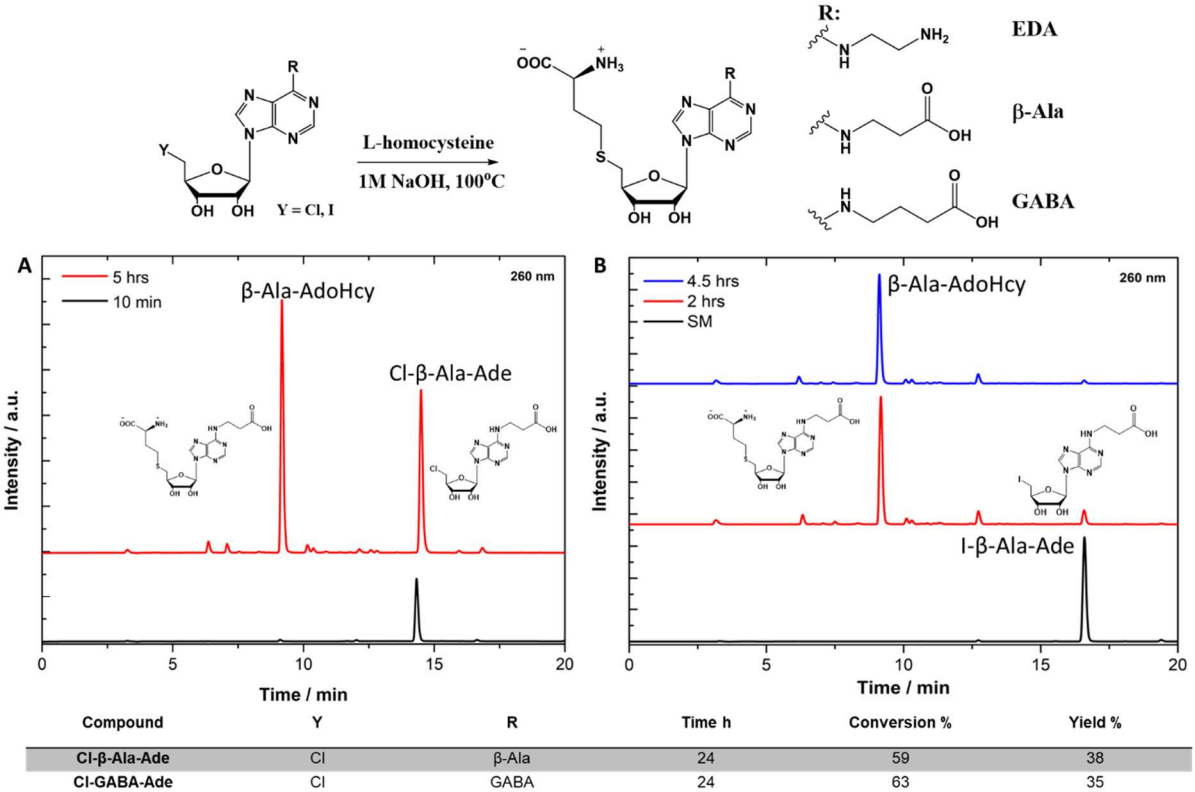


Fig. 35 Reaction scheme of the coupling of 5'-Cl and 5'-I activated functionalised adenosine analogues with L-homocysteine

The coupling was performed with 2 equivalents of L-homocysteine and 3 equivalents of 1M NaOH at 100°C. All chlorinated compounds showed significantly lower conversion than iodinated ones. For Cl-β-Ala-Ade as a starting material the conversion after 5 hours reached 57% (Fig. 36 A) whereas I-β-Ala-Ade showed conversion of 90% just after 2 hours reaching 95% within next 2.5 hours (Fig. 36 B). As shown in the table in Fig. 36, the HPLC conversions of all three reacted chlorinated reagents after 24 hours are significantly lower (up to 36 %) than those in the reaction of iodinated nucleosides after 4.5 hours (I-β-Ala-Ade and I-GABA-Ade) or 10 minutes (I-EDA-Ade). For all three iodinated compounds that were reacted, the reaction can be quenched just after 2 hours if necessary as almost complete conversion is reached. Such an improvement in the reaction rate may help to avoid degradation of any sensitive N6-modification which may have been introduced beforehand.



Compound	Y	R	Time h	Conversion %	Yield %
Cl-β-Ala-Ade	Cl	β-Ala	24	59	38
Cl-GABA-Ade	Cl	GABA	24	63	35
Cl-EDA-Ade	Cl	EDA	5	62	37
I-β-Ala-Ade	I	β-Ala	4.5	95	81
I-GABA-Ade	I	GABA	4.5	97	74
I-EDA-Ade	I	EDA	0.17	98	67

Fig. 36 Reactivity of chlorinated and iodinated functionalized nucleosides in the coupling reaction with L-homocysteine. (A) HPLC profiles of the coupling reaction of Cl-β-Ala-Ade with L-homocysteine after 10 min and 5 hours; HPLC conditions: 20 mM ammonium formate (pH = 3.5)/MeCN, gradient 3-100% MeCN over 60 min.(B) HPLC profiles of the coupling reaction of I-β-Ala-Ade with L-homocysteine at the start of the reaction, after 2 hours and 4.5 hours. Table: Comparison of coupling reaction HCPL conversions and yields for chlorides and iodides of three functionalised adenosine derivatives.

Having showed that functionalized nucleosides bearing 5'-iodine are better precursors of AdoHcy analogues, we coupled L-homocysteine with other analogues bearing various modifications in the nucleobase. Hence, a library of N^6-substituted AdoHcy analogues bearing terminal amine, carboxyl, hydroxyl, azide alkene and alkyne groups was produced (Fig. 37).

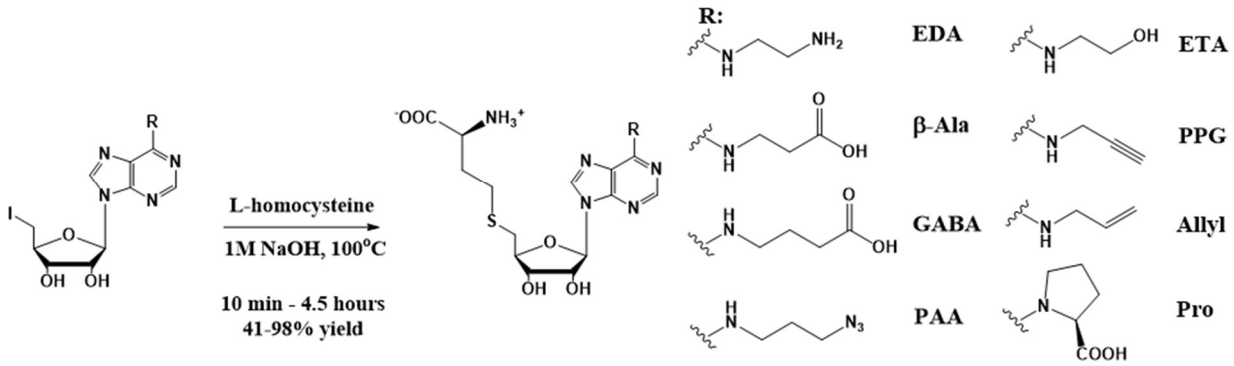


Fig. 37 Synthesis of N^6-substituted AdoHcy analogues.

Lastly, to show that newly synthesised AdoHcy analogues can be further transformed into AdoMet analogues in the alkylation reaction, we reacted two of the synthesised AdoHcy analogues with methyl iodide (See II.1.2.6).

For synthesized compounds bearing L-proline as a modification we observed interesting features in the NMR spectra, which will be discussed in the following paragraph.

II.1.2.5. NMR analysis of L-proline modified analogues

We found that NMR spectra of modified 5'-iodoadenosine with L-proline show an interesting feature. Two protons in the adenine ring (H2 and H8) appear as four singlets in ^{1}H NMR spectrum instead of two singlets (Fig. 38 A). The integration of these four peaks shows that they originate from two protons. ^{1}H-^{13}C HSQC NMR spectrum in this area shows that two singlets of carbons C2 and C8 are also split (Fig. 38 B). In addition

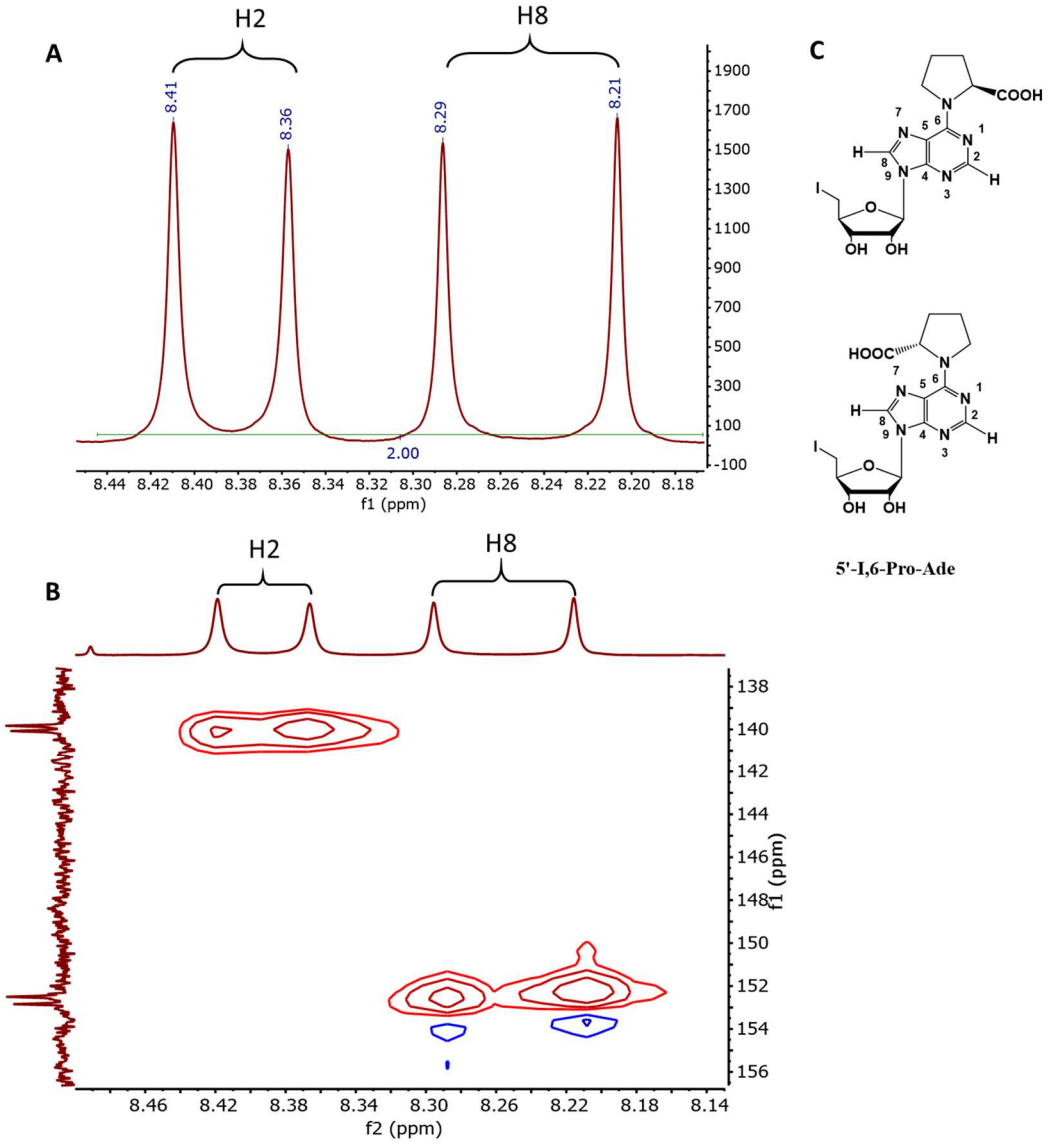


Fig. 38 NMR signals of two protons (H2 and H8) in the adenine of 5'-I,6-Pro-Ade ring in (A) ^{1}H NMR spectrum and (B) ^{1}H-^{13}C HSQC spectrum. (C) Proposed rotamers of 5'-I,6-Pro-Ade and two possible positions of -COOH of L-proline affecting chemical shift of both H2 and H8 as well as C2 and C8 in NMR spectra. Spectra taken in DMSO.

each pair of proton peaks is coupled to a pair of peaks of relevant carbon. Such split of both proton and carbon peaks suggests that there are two isomers of I-Pro-Ade present. Due to the fact that proline can possibly rotate along the C-N bond we suggest two possible rotamers (Fig. 38 C). A carboxyl group of the L-proline in each isomer have a different influence on C2/H2 and C8/H8 and hence their different chemical shifts.

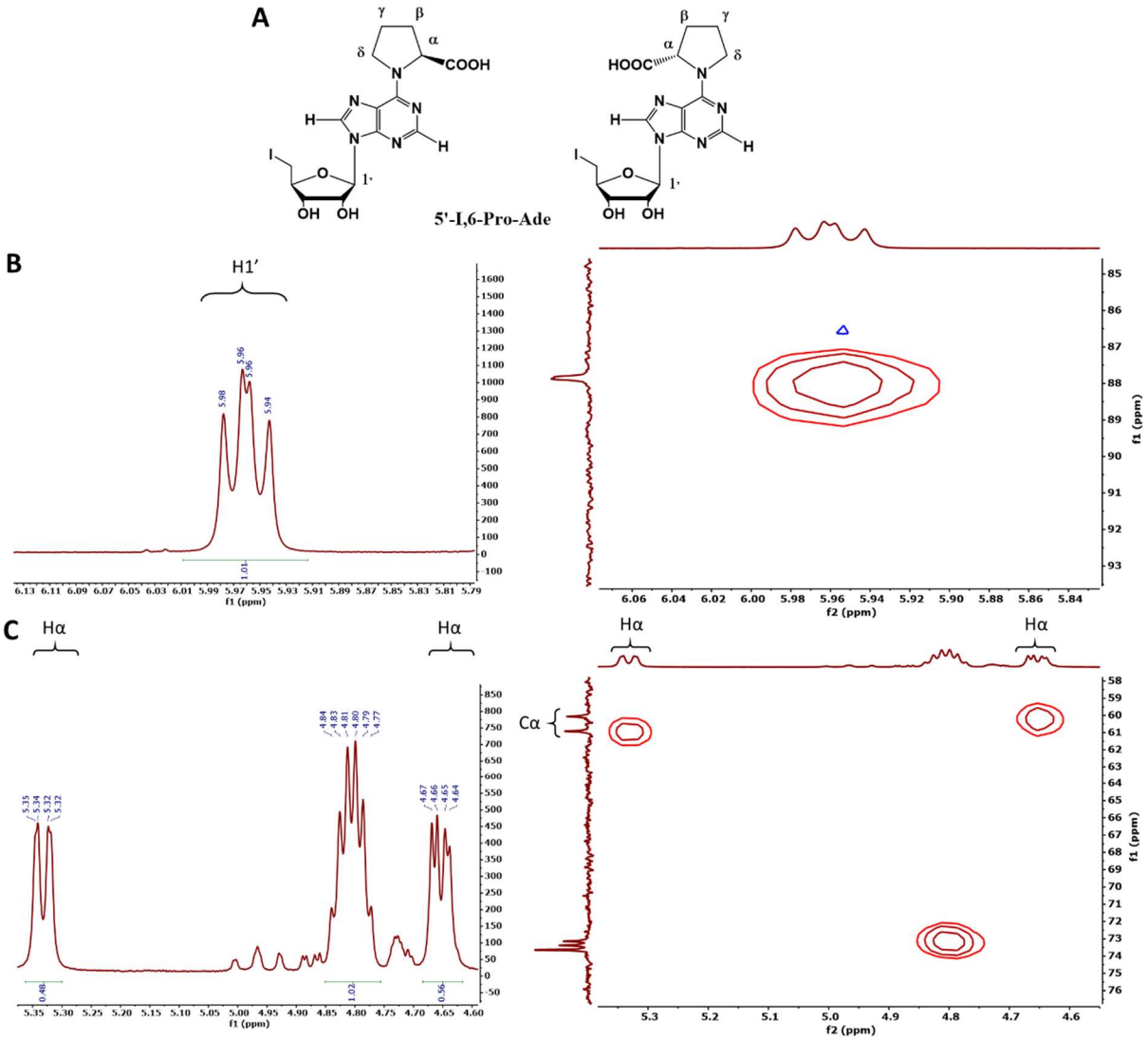


Fig. 39 Proposed rotamers of 5'-I,6-Pro-Ade and two possible positions of -COOH of L-proline (**A**) affecting a peak of H1' in ribose (**B**) and Hα in L-proline (**C**) ^{1}H NMR spectrum (left panel) and ^{1}H-^{13}C HSQC spectrum (right panel) are shown. Spectra taken in DMSO.

In addition to the split of two protons in adenine ring we observed that NMR signals of a proton attached to a carbon adjacent to carboxyl group (Hα) in L-proline and H1' proton in ribose (Fig. 39 A) are split. A signal of H1' is observed as two overlapping dublets at 5.96 ppm with 1.01 integration (Fig. 39 B) and Hα signal is split into two distinct peaks at 5.33 ppm and 4.66 ppm with a sum of integrations 1.04 (Fig. 39 C). Signals of carbons attached to these two protons are split which can be observed in ^{1}H-^{13}C HSQC spectrum. A large split of signal of Hα suggests further that these two compounds are rotamers with L-proline locked in one of two positions (Fig. 39 A).

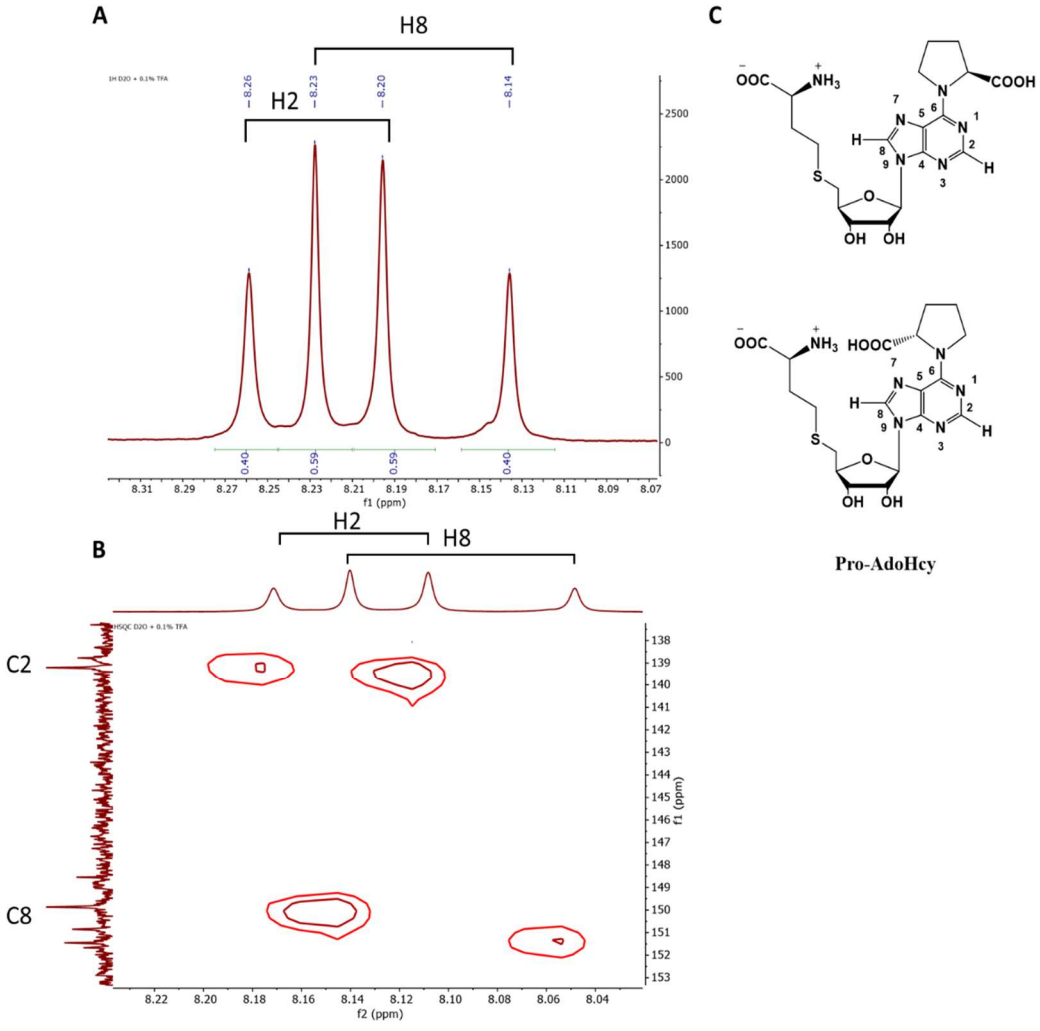


Fig. 40 NMR signals of two protons (H2 and H8) in the adenine of Pro-AdoHcy ring in (A) ^{1}H NMR spectrum and (B) ^{1}H-^{13}C HSQC spectrum. (C) Proposed rotamers of Pro-AdoHcy and two possible positions of -COOH of L-proline affecting chemical shift of both H2 and H8 as well as C2 and C8 in NMR spectra. Spectra taken in D_2O+0.1% TFA.

The evidence of two rotamers is also present in NMR of L-proline modified AdoHcy analogue (Pro-AdoHcy - Fig. 40 C). Similarly to the 5'-I,6-Pro-Ade NMR, singlets originated from H2 and H8 are split (Fig. 40 A). An integration ratio for each pair is about 60:40 which suggests that one of the rotamers is favourable over the other. The ^{1}H-^{13}C HSQC spectrum also shows corresponding split of C2 and C8 signals.

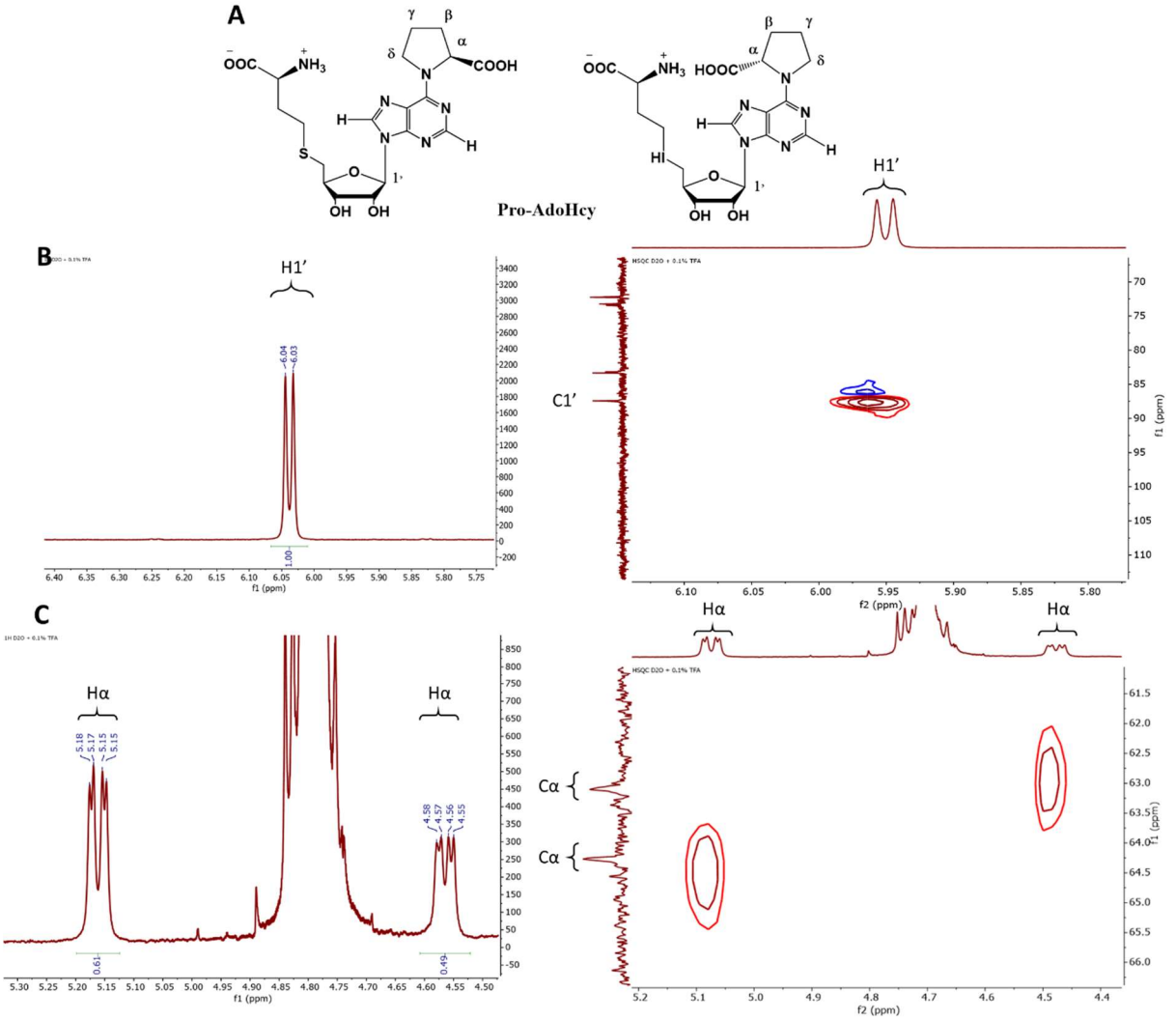


Fig. 41 Proposed rotamers of Pro-AdoHcy and two possible positions of -COOH of L-proline (A) with no effect on a peak of H1' in ribose (B) and affecting Hα in L-proline (C) 1H NMR spectrum (left panel) and 1H-13C HSQC spectrum (right panel) are shown. NMR spectra taken in D_2O+0.1% TFA.

Similarly to 5'-I,6-Pro-Ade, a signal of Hα in L-proline is split with about 60:40 ratio which can also be observed in ^{1}H-^{13}C HSQC spectrum (Fig. 41 C). In contrast, a singal of H1' proton in ribose is unaffected by the presence of the rotamers (Fig. 41 B).

This formation of two different conformational isomers could be possibly exploited in biological applications where usually one isomer of a biologically active compounds is bound preferentially by an enzyme. However, at this stage both rotamers are present in the solution and likely can dynamically switch their conformation. Hence, if the formation of particular rotamer in controllable manner in a solution or both could be separately purified, such analogues could be used in the studies of a dynamic of enzymatic reactions. However, a possible method of controllable formation of each rotamer was not pursued further in this work.

II.1.2.6. Synthesis of AdoMet analogues

Doubly-activated AdoMet analogues are most often synthesised in the alkylation reaction of the sulphur in AdoHcy molecule. In this reaction, an activated alkyl precursor, usually idodie, bromide or tosyl, is reacted under acidic conditions e.g. 1:1 mixture of formic and acetic acids, with the AdoHcy molecule (Fig. 42). Nucleophilic attack occurs at sulphur rather than amine due to acidic conditions which protonates the amines and hence decreases their nucleophilic character. When the alkyl halide is used, silver perchlorate is usually added to the rection mixture which acts as a Lewis acid and facilitates the reaction. As a result of this reaction two isomers R,S and S,S of AdoMet analogue are created which, in some cases, can be separated during purification by HPLC. This reaction has been widely applied to create AdoMet analogues bearing a wide range of functional groups in place of a methyl group.[151, 160]

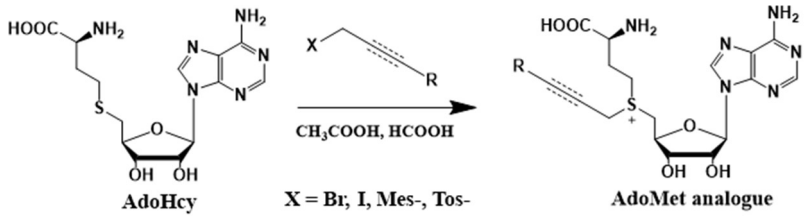


Fig. 42 Synthesis of doubly-activated AdoMet analogues via the alkylation reaction.

Therefore, to show that functionalised AdoHcy analogues can be further reacted to obtain AdoMet analogues. We chose AdoHcy analogues bearing amine (EDA-AdoHcy) and carboxyl (GABA-AdoHcy) groups that could potentially react in the alkylation reaction (Fig. 43). As an alkylation reagent we used a reactive primary halide – methyl iodide to obtain AdoMet analogues. Amines and carboxylic acid can undergo a substitution reaction with methyl iodide under basic conditions. Even though the AdoMet synthesis is carried out in the presence of acid, we decided to test whether the terminal amine and carboxyl groups in the introduced moieties were modified during the alkylation reaction.

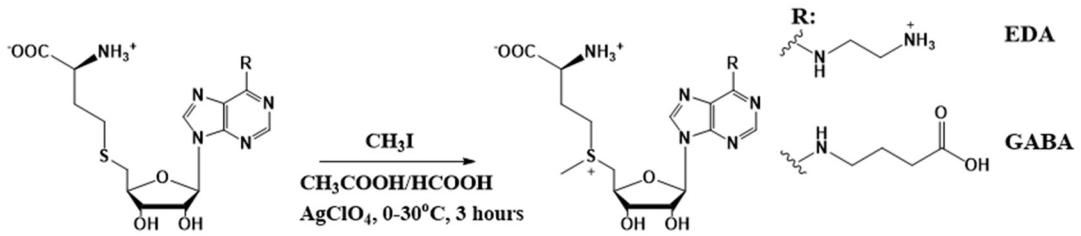


Fig. 43 Alkylation of N^6-substituted AdoHcy analogues with methyl iodide.

HPLC profiles of the EDA-AdoMet synthesis reaction show peak of starting material (EDA-AdoHcy) and peak at 9.8 min and a product at 3.7 min (Fig. 44). Mass spectrometry analysis showed a mass of 442.19 Da which indicates an addition of a single methyl group to EDA-AdoHcy. The fact that this compound elutes significantly

earlier than starting material is consistent with its increased polarity. If an amine group was methylated the resulting compound would be less polar hence should elute after the starting material. In addition, the mass spectrum shows a common peak of a mass spectrometry product with a loss of homocysteine (341.14 Da). However, this does not resolve the position of the methyl group. However, a peak that can be assigned to the

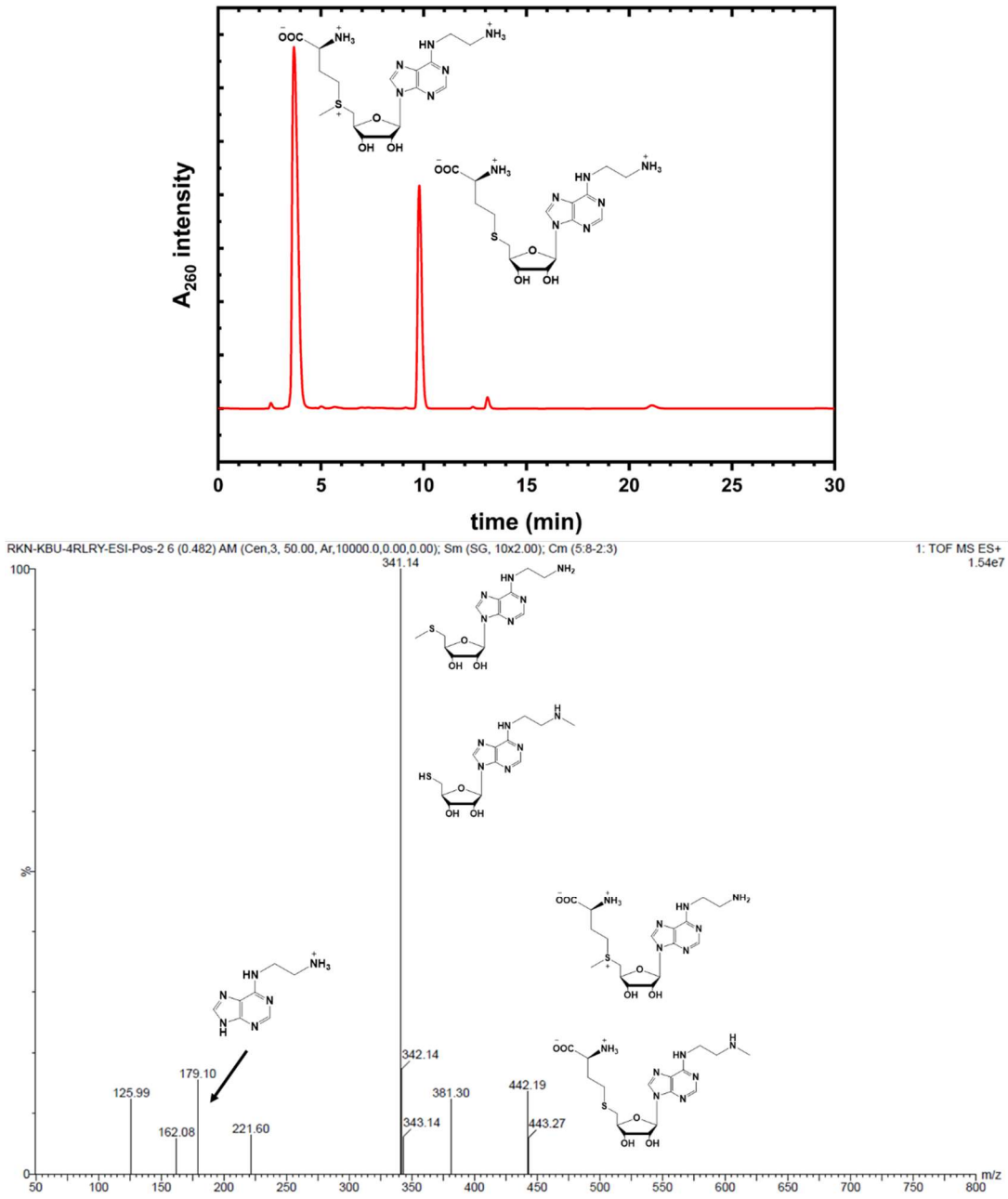


Fig. 44 HPLC profiles of EDA-AdoMet (top). Elution conditions: 20 mM ammonium formate pH 3.5/MeOH gradient 3-75% MeOH. Mass spectrum of both peaks (3.7 min) from EDA-AdoMet synthesis HPLC profile.

ethylenediamine functionalised adenine is present (179.10 Da) whereas a peak of possible methylation of ethylenediamine is not present. This is indicative of a methyl group being attached to the sulphur centre rather than an amine group in the ethylenediamine modification.

Similarly, the HPLC profile of the GABA-AdoMet synthesis reaction shows peak of starting material at 18.2 min. In addition two poorly resolved peaks can be observed at 11.1 and 11.4 min which represent two isomers of GABA-AdoMet. Due to the poor separation both peaks were analysed on mass spectrometry together. The mass spectrum shows a major peak of mass 485.18 Da from addition of a single methyl group to the GABA-AdoHcy (Fig. 45). Moreover, a peak of 336.13 Da can be distinguished which can be assigned to GABA-AdoMet with a loss of methionine. Lower retention time of both peaks suggests that it is a desired product, GABA-AdoMet, rather than methylation of a terminal carboxyl group in the 6 position. A small peak at 499.20 Da is present which corresponds to the doubly methylated product. However, peaks resulting from loss of either homocysteine (398.15 Da) or methionine (350.15 Da) are not present. This suggest that if a doubly methylated AdoMet analogue is present in the analysed fraction, its content is rather low.

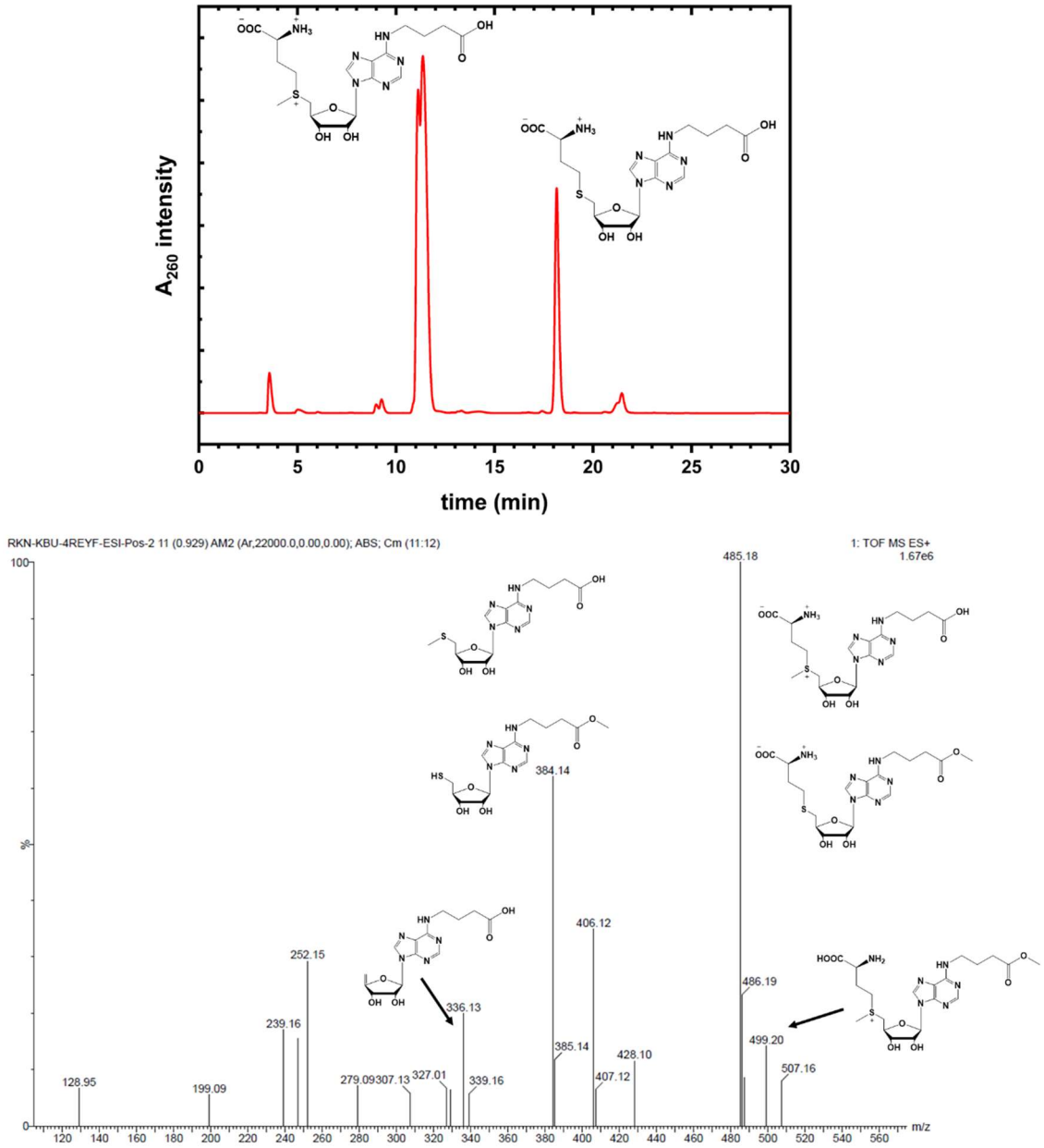


Fig. 45 HPLC profiles of GABA-AdoMet (top). Elution conditions: 20 mM ammonium formate pH 3.5/MeOH gradient 3-75% MeOH. Mass spectrum of both peaks (11.1 and 11.4 min) from GABA-AdoMet synthesis HPLC profile.

As anticipated, the terminal amino group in EDA-AdoHcy is protonated and hence loses its nucleophilic properties and does not undergo methylation under the present reaction conditions. GABA-AdoMet was also successfully synthesized with only desired product of a sulphur methylation present over a methylation of GABA. However, mass spectrometry analysis of the fraction collected suggests a minor

presence of a doubly methylated by-product. During the purification on preparative RP-HPLC the GABA-AdoMet and EDA-AdoMet analogues were collected as a mixture of two isomers.

In order to assess the activity of both synthesized AdoMet analogues as cofactors of MTase enzyme we tested their activity against M.TaqI enzyme.

II.1.2.7. The activity of N^6-substituted AdoMet analogues as M.TaqI cofactors

The ability of both synthesized AdoMet analogues to act as an effective cofactor in the MTase-mediated transalkylation reaction was tested in a restriction assay adapted from the literature.[222] In such an assay, DNA is labelled with the cofactor and methyltransferase of choice, followed by restriction with a complimentary restriction enzyme. In our assay we used a bacterial M.TaqI methyltransferase that recognises a palindromic 5'-TCGA-3' sequence and alkylates the N6 residue of adenine within the

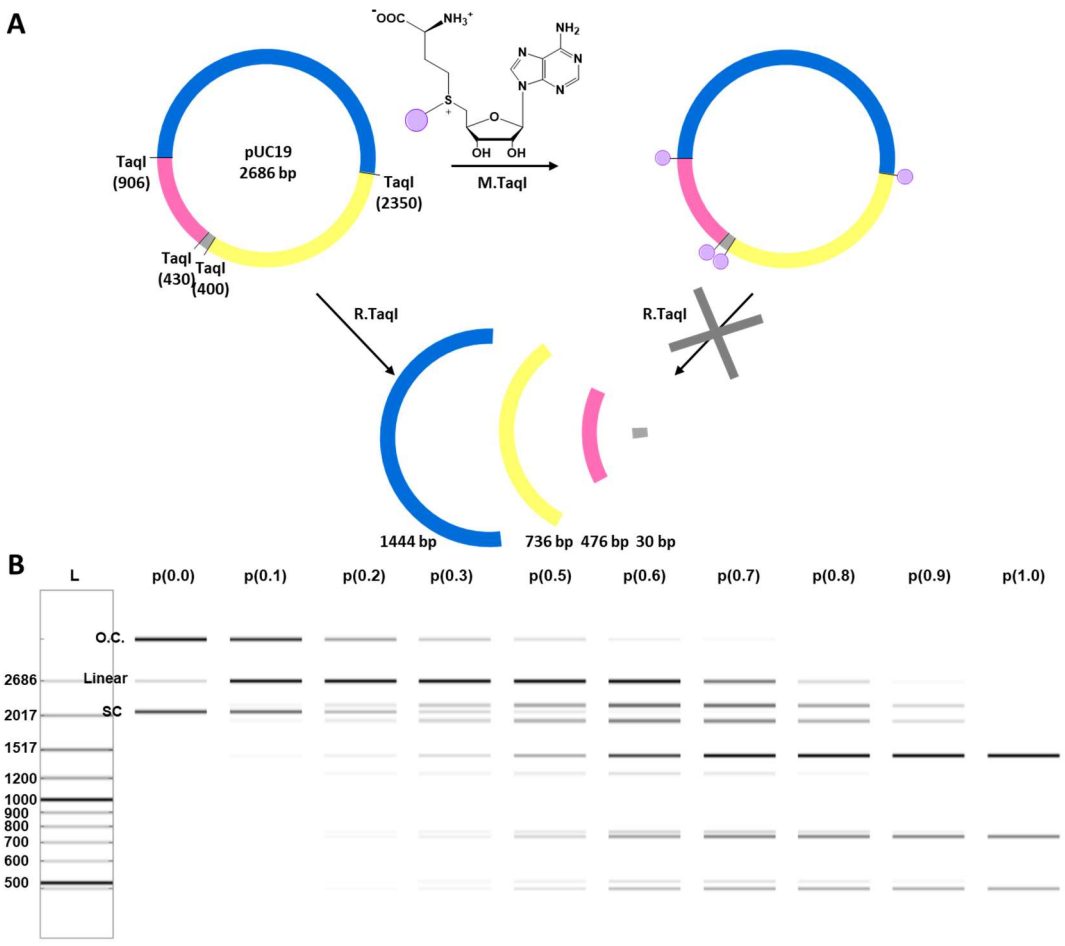


Fig. 46 (**A**) Schematic representation of pUC19 DNA plasmid transalkylation by M.TaqI and AdoMet analogue cofactor followed by restriction with complementary restriction enzyme R.TaqI. Upon successful transalkylation reaction R.TaqI shows no activity and hence the plasmid is intact. The unlabelled plasmid DNA is cut up in to 4 fragments of 1444 bp, 736 bp, 476 bp and 30 bp sizes. (**B**) Computationally modelled a pattern of bands of gels depending on the increasing efficiency of digestion by R.TaqI. p(0.0) – unsuccessful digestion, pUC19 shows three bands of open circular (O.C.), linear and supercoiled (SC) form; p(1.0) – complete digestion; three fragments of 1444 bp, 736 bp and 476 bp are visible.

5’-TCGA-3’ sequence in the presence of a cofactor. The complimentary endonuclease (TaqαI or R.TaqI) recognises the same sequence as M.TaqI and cuts DNA if adenine within this sequence is unmodified. If the adenine within 5’-TCGA-3’ sequence is alkylated this part of DNA is left intact (Fig. 46 A).

The efficiency of the transalkylation reaction can be then monitored by gel electrophoresis where digested DNA can be distinguished from the intact DNA molecule. As a DNA template we used a 2686 bp long circular plasmid DNA – pUC19. It contains four available 5’-TCGA-3’ sites and alkylation can occur on each strand of the DNA, so a total of eight alkylation events per DNA molecule are possible. The location of all sites is distributed in such a way that upon a complete digestion by R.TaqI enzyme, pUC19 is cut into four distinctive in size DNA fragments, which facilitates quantitative analysis in the gel.

The pattern of DNA bands (fragments) in the gel, post electrophoresis, changes with the efficiency of the initial M.TaqI labelling and can be computationally modelled allowing an estimation of the labelling efficiency (Fig. 46 B). Upon complete alkylation (no digestion), there are three possible forms of the pUC19 plasmid identified – open circular (O.C.), linear and supercoiled (SC) with O.C. and SC being dominant forms. Following complete restriction of the plasmid with R.TaqI enzyme, three distinct bands can be observed representing 1444 bp, 736 bp and 476 bp fragments. The shortest 30 bp long fragment is usually not observed in the gel.

The activity of EDA-AdoMet and GABA-AdoMet (mixtures of isomers) was probed in the restriction assay (Fig. 47). Gels for both cofactors show that at the highest concentration of the M.TaqI enzyme (1.9 ng/μL) the pUC19 plasmid is completely

methylated. The efficiency of the labelling slowly decreases reaching about 70% for EDA-AdoMet and 30-40% for GABA-AdoMet at the lowest concentration of the M.TaqI

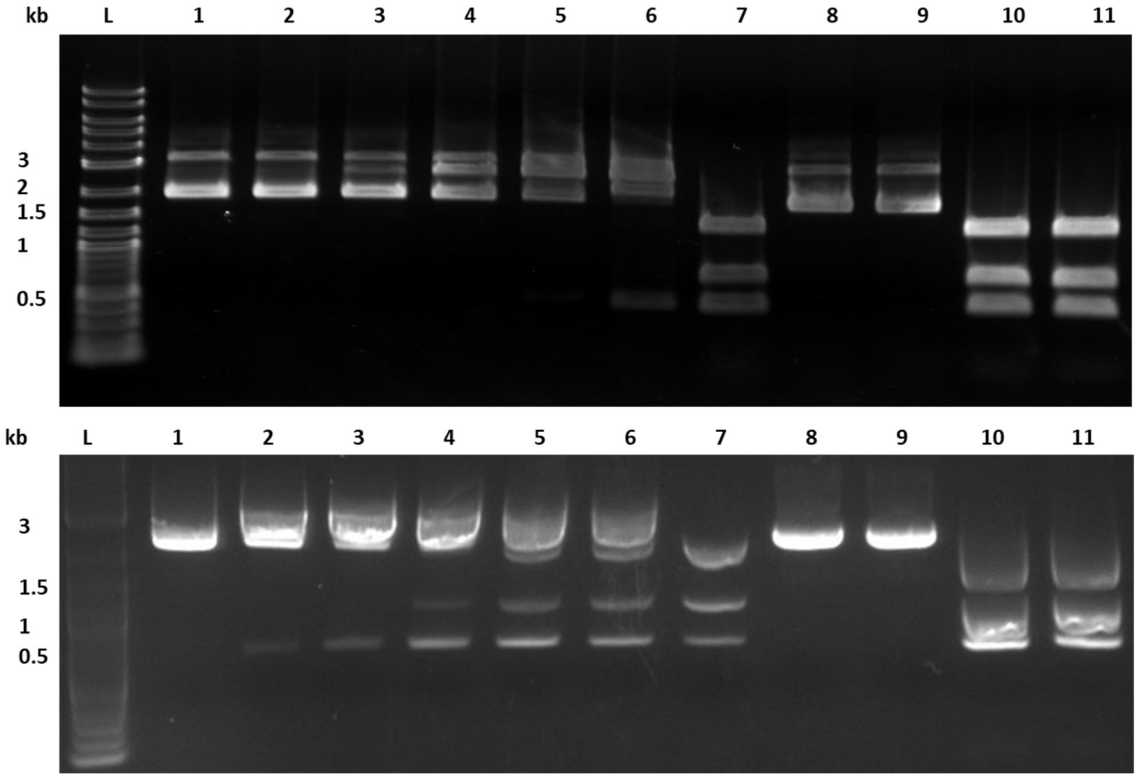


Fig. 47 Restriction assay of EDA-AdoMet (top) and GABA-AdoMet (bottom) with M.TaqI enzyme and pUC19 plasmid DNA. (Top) EDA-AdoMet concentration at 250 μM, Lanes 1-6: serial dilutions of M.TaqI enzyme (1.9 - 0.06 ng/μl); 7: control of restriction enzyme – pUC19 fully digested in presence of EDA-AdoMet (250 μM); 8-9: positive control with AdoMet of complete protection; 10: negative control with no cofactor; 11: negative control with no M.TaqI enzyme; (Bottom) GABA-AdoMet concentration at 250 μM, Lanes 1-6: serial dilutions of M.TaqI enzyme (1.9 -0.06 ng/μl); 7: control of restriction enzyme – pUC19 fully digested in presence of GABA-AdoMet (250 μM); 8-9: positive control with AdoMet of complete protection; 10: negative control with no cofactor; 11: negative control with no M.TaqI enzyme

(0.06 ng/μL). Hence, both cofactors are recognised by M.TaqI methyltransferase enzyme and act as active methyl group donors in transalkylation reaction. EDA-AdoMet is more active at low concentrations of M.TaqI enzyme like due to a smaller modification introduced in the 6 position of adenine. In addition successful protection of pUC19 with a mixture of isomers shows that the presence of inactive isomer does not hinder the M.TaqI activity.

II.1.3. Conclusions

In conclusion, we have developed an efficient, versatile and protecting-group-free synthesis of N^6-substituted AdoHcy/AdoMet analogues. We have developed a new doubly-activated intermediate that allows a selective introduction of a wide range of functional moieties such as primary amines, carboxyl groups, azides, alkynes and alkenes in the N6 position of adenine in AdoHcy molecule. The presence of iodine in 5' position of ribose significantly improves the time of the crucial step of coupling with L-homocysteine, from more than 24 hours to less than 5 hours and to as little as 10 minutes for some compounds. In addition we showed that AdoHcy analogues bearing nucleophilic groups in the introduced moiety can be successfully used in the synthesis of corresponding AdoMet analogues in the alkylation reaction.

In addition, L-proline modified AdoHcy analogues show an interesting behaviour. During the NMR analysis, two rotamers of 5'-I,6-Pro-Ade and Pro-AdoHcy were identified. These could possibly have different biological activity if applied with AdoHcy recognising enzymes, especially if the purification of each rotamer separately or a method to interconvert them/stabilise would be available.

Both synthesised AdoMet analogues bearing polar groups in the 6 position of adenine (EDA-AdoMet, GABA-AdoMet) are recognised by bacterial M.TaqI methyltrasferase enzyme and act as donors of a methyl group. This shows a potential for AdoMet analogues modified with polar groups to be effective cofactors, however studies applying a range of modifications as well as methyltransferase are required.

We believe that this approach will widen a range of possible modifications to the AdoHcy/AdoMet structure. It will help in the synthesis of versatile tools in studying of

epigenetic regulation, design of AdoHcy-structure based MTases inhibitors or capture compounds.

II.1.4. Materials and methods

II.1.4.1. General materials

All chemicals were purchased from Sigma-Aldrich except 6-chloropurine riboside and L-homocysteine which were purchased from Fluorochem, thionyl chloride purchased from Acros Organics, L-proline purchased from Thermo Fisher Scientific and allylamine and propargylamine purchased from Alfa Aesar. CutSmart® buffer, pUC19 plasmid, proteinase K, TaqαI (R.TaqI) were purchased from New England BioLabs (NEB). NMR data was acquired on a Bruker Avance III operating at 400 MHz and 500 MHz. HRMS spectra were obtained on a Xevo® G2-XS-ToF (Waters) and Synapt-G2-S from electrospray ionization (ESI) and time-of-flight (TOF) measurement in negative or positive ion mode.

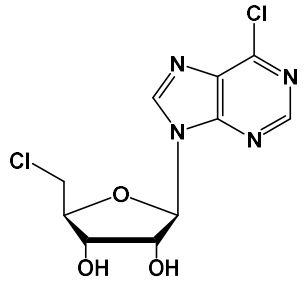
II.1.4.2. HPLC analysis, concentration and yield determination

Analytical RP-HPLC was performed on Shimadzu LC-20 Prominence equipped with ACE 5 C18 (250 x 4.6 mm, flow rate 1mL/min). Preparative RP-HPLC was performed on Agilent Technologies 1260 Infinity equipped with ACE 5 C18 (250 x 21.2 mm, 100 Å, flow rate 10 mL/min). All HPLC profiles were acquired at 260 nm.

The concentration and yields for N^6-modified AdoMet analogues were determined by UV absorbance measurements at 260 nm, performed in 0.1 % acetic acid. For the calculation, molar extinction coefficient ε_{260} = 15400 $dm^3\ mol^{-1}\ cm^{-1}$ was used.

II.1.4.3. Synthesis procedures

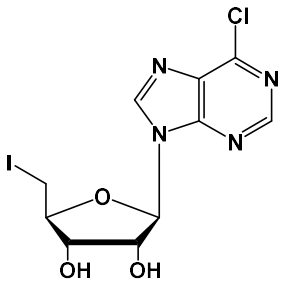
Synthesis of (2R,3R,4S,5S)-2-(6-chloro-9H-purin-9-yl)-5-(chloromethyl)tetrahydrofuran-3,4-diol (5',6-diCl-Ade)



To a cold suspension of 6-chloropurine riboside (1 g, 3.5 mmol) in acetonitrile (10 mL), distilled thionyl chloride (0.76 mL, 3 eq) was added. Then pyridine (0.56 mL, 2 eq) was added and the reaction was stirred at 0°C for 4 hours and overnight at room temperature. Then, the solvent was removed under reduces pressure and the sample was dissolved in 20 mL of MeOH. 1 mL of water and 2 mL of 35% aqueous solution of ammonia was added. The reaction was stirred for 3 hours. After 1 hour additional 0.6 mL of ammonia was added. The solvent was removed under reduced pressure, 25 mL of 5% citric acid was added and product was extracted with ethyl acetate. The organic layer was washed with $NaHCO_3$, brine and dried over anhydrous Na_2SO_4. The solvent was removed under reduced pressure yielding a yellow solid (853.4 mg, 80%): ^{1}H NMR (400 MHz, DMSO-d6) δ 8.92 (s, 1H, 8-*H*), 8.83 (s, 1H, 2-*H*), 6.08 (d, J = 5.3 Hz, 1H, 1'-*H*), 5.72 (br. s, 1H, 2'-O*H*), 5.56 (br. s, 1H, 3'-O*H*), 4.80 – 4.73 (m, 1H, 2'-*H*), 4.31 – 4.23 (m, 1H, 3'-*H*), 4.16 (ddd, J = 6.3, 4.9, 4.2 Hz, 1H, 4'-*H*), 3.97 (dd, J = 11.7, 4.9 Hz, 1H, 5'-*H*), 3.88 (dd, J = 11.7, 6.3 Hz, 1H, 5''-*H*); ^{13}C NMR (101 MHz, DMSO) δ 151.92 (6-C), 151.65 (2-C), 149.48 (4-C), 146.10 (8-C), 131.44 (5-C), 88.17 (1'-C), 83.98 (4'-

C), 72.93 (2'-C), 71.14 (3'-C), 44.71 (5'-C); TOF MS ES (-) m/z $[M+Cl]^-$ calcd: 338.9818 found: 338.9815.

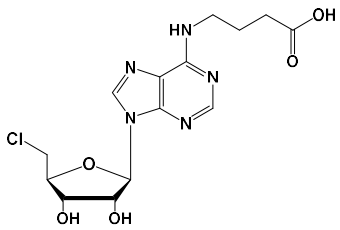
Synthesis of (2R,3R,4S,5S)-2-(6-chloro-9H-purin-9-yl)-5-(iodomethyl)tetrahydrofuran-3,4-diol (5'-I,6-Cl-Ade)



To a cold solution of 6-chloropurine riboside (500 mg, 1.75 mmol) in NMP (3 mL), imidazole (773.5 mg, 6.5 eq) and triphenylphosphine (1516 mg, 3.3 eq) were added. Then solution of iodine (1464 mg, 3.3 eq.) in 2 mL of NMP was added dropwise and the reaction was stirred overnight at room temperature. Then, 10 mL of water was added and product was extracted with ethyl acetate. The organic layer was washed with brine and dried over anhydrous Na_2SO_4. The solvent was removed under reduced pressure, the resulting yellow oil was stored overnight in the fridge. Crystallized triphenylphosphine oxide was filtered off and the crude was purified by preparative RP-HPLC (50 – 100% MeOH in water over 60 minutes). Collected fraction was lyophilized yielding a white solid (526.8 mg, 76%): ^{1}H NMR (400 MHz, DMSO-d6) δ 8.94 (s, 1H, 8-*H*), 8.84 (s, 1H, 2-*H*), 6.07 (d, J = 5.5 Hz, 1H, 1'-*H*), 5.69 (d, J = 5.7 Hz, 1H, 2'-O*H*), 5.55 (d, J = 5.2 Hz, 1H, 3'-O*H*), 4.82 (ddd, J = 5.7, 5.1 Hz, 1H, 2'-*H*), 4.21 (ddd, J = 5.2, 3.8 Hz, 1H, 3'-*H*), 4.04 (ddd, J = 7.0, 5.8, 3.8 Hz, 1H, 4'-*H*), 3.62 (dd, J = 10.5, 5.8 Hz, 1H, 5'-*H*), 3.49 (dd, J = 10.5, 7.0 Hz, 1H, 5''-*H*); ^{13}C NMR (101 MHz, DMSO) δ

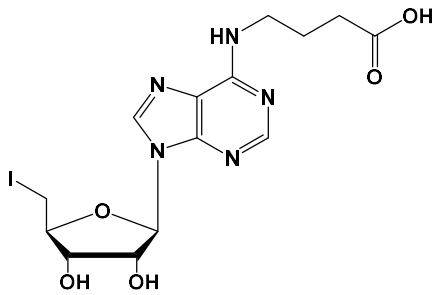
151.64 (2-*C*), 145.99 (8-*C*), 87.91 (1'-*C*), 83.92 (4'-*C*), 72.84 (2'-*C*), 72.75 (3'-*C*), 7.23 (5'-*C*). TOF MS ES (+) $[M+H]^+$ calcd: 396.9559, found: 396.9561

Synthesis of 4-((9-((2R,3R,4S,5S)-5-(chloromethyl)-3,4-dihydroxytetrahydrofuran-2-yl)-9H-purin-6-yl)amino)butanoic acid (5'-Cl,6-GABA-Ade)



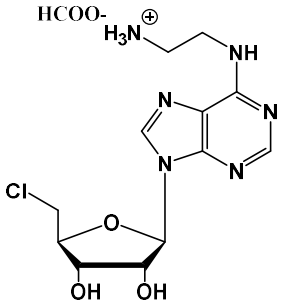
To a solution of 5',6-diCl-Ade (300 mg, 0.99 mmol) in MeOH (3 mL), GABA (305.91 mg, 3 eq) in 1 mL of water was added. Then triethylamine (821 µL, 6 eq.) was added and the reaction was stirred for 8 hours. The MeOH was removed under reduced pressure and pH was adjusted to 3 with 1M HCl. The precipitate was collected and washed with cold water and dried over P_2O_5 yielding an off-white solid and used in the next step with no further purification (264.5 mg, 72%): 1H NMR (400 MHz, DMSO-d6) δ 12.03 (s, 1H, COO*H*), 8.34 (s, 1H, 2-*H*), 8.22 (s, 1H, 8-*H*), 7.91 (s, 1H, N*H*), 5.94 (d, J = 5.6 Hz, 1H, 1'-*H*), 5.60 (s, 1H, 2'-O*H*), 5.46 (s, 1H, 3'-O*H*), 4.76 (dd, J = 5.9, 4.8 Hz, 1H, 2'-*H*), 4.23 (dd, J = 4.8, 3.9 Hz, 1H, 3'-*H*), 4.09 (ddd, J = 6.4, 5.1, 3.9 Hz, 1H, 4'-*H*), 3.95 (dd, J = 11.6, 5.1 Hz, 1H, 5'-*H*), 3.84 (dd, J = 11.6, 6.4 Hz, 1H, 5''-*H*), 3.49 (s, 2H, NHCH_2CH$_2$), 2.27 (t, J = 7.4 Hz, 2H, CH$_2$CH_2COOH), 1.82 (tt, J = 7.4, 7.2 Hz, 2H, CH$_2$CH_2CH$_2$); 13C NMR (101 MHz, DMSO) δ 152.40 (8-*C*), 139.29 (2-*C*), 87.22 (1'-*C*), 83.42 (4'-*C*), 72.40 (2'-*C*), 71.01 (3'-*C*), 44.56 (5'-*C*), 38.82 (NH*C*H$_2$CH$_2$), 30.90 (CH$_2$*C*H$_2$COOH), 24.27 (CH$_2$*C*H$_2$CH$_2$). TOF MS ES (-) m/z $[M-H]^-$ calcd: 370.0918 found: 370.0916.

Synthesis of 4-((9-((2R,3R,4S,5S)-5-(iodomethyl)-3,4-dihydroxytetrahydrofuran-2-yl)-9H-purin-6-yl)amino)butanoic acid (5'-I,6-GABA-Ade)



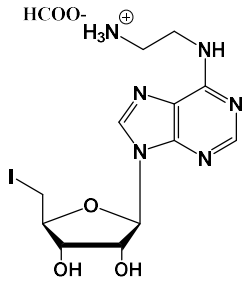
To a solution of 5'-I,6-Cl-Ade (150 mg, 0.38 mmol) in MeOH (1.5 mL), GABA (117 mg, 3 eq) in 200 µL of water was added. Then triethylamine (317 µL, 6 eq.) was added and the reaction was stirred for 8 hours. The MeOH was removed under reduced pressure and pH was adjusted to 4-5 with 1M HCl. The precipitate was collected and washed with cold water and dried over P_2O_5 yielding a white solid and used in the next step with no further purification (129 mg (95% purity with traces of GABA and TEA), 70%): 1H NMR (400 MHz, DMSO-d6) δ 11.98 (s, 1H, COO*H*), 8.37 (s, 1H, 2-*H*), 8.22 (s, 1H, 8-*H*), 7.91 (s, 1H, N*H*CH_2), 5.93 (d, J = 5.7 Hz, 1H, 1'-*H*), 5.72 – 5.33 (m, 2H, 2'-O*H,* 3'-O*H*), 4.81 (dd, J = 5.7, 5.1 Hz, 1H, 2'-*H*), 4.20 – 4.16 (m, 1H, 3'-*H*), 4.04 – 3.95 (m, 1H, 4'-*H*), 3.66 – 3.57 (m, 1H, 5'-*H*), 3.56 – 3.41 (m, 3H, 5''-*H*, NHC*H*$_2$$CH_2$), 2.28 (t, J = 7.3 Hz, 2H, CH_2C*H*$_2$COOH), 1.82 (tt, J = 7.3, 7.2 Hz, 2H, CH_2C*H*$_2$$CH_2$). 13C NMR (101 MHz, DMSO-H6) δ 174.29 (COOH), 152.68 (8-*C*), 139.72 (2-*C*), 87.53 (1'-*C*), 83.91 (4'-*C*), 73.19 (3'-*C*), 72.78 (2'-*C*), 40.15 (5'-*C*, 5''-*C*, NH*C*H_2, under DMSO peak), 31.17 (CH_2*C*H_2COOH), 24.52 (CH_2*C*$H_2$$CH_2$). TOF MS ES (-) m/z $[M-H]^-$ calcd: 462.0274; found: 462.0284.

Synthesis of 2-((9-((2R,3R,4S,5S)-3,4-dihydroxy-5-(chloromethyl)tetrahydrofuran-2-yl)-9H-purin-6-yl)amino)ethan-1-aminium formate (5'-Cl,6-EDA-Ade)



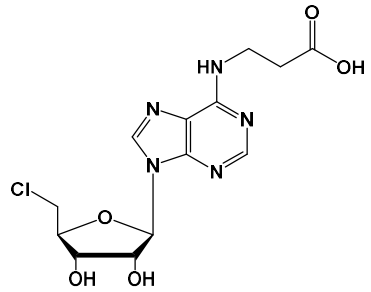
To a solution of 5',6-diCl-Ade (150 mg, 0.38 mmol) in MeOH (1.5 mL), EDA (152 µL, 6 eq) was added and reaction was stirred for 2 hours. The MeOH was removed under reduced pressure and pH was adjusted to 3 with 1M HCl. Resulting solution was purified by preparative RP-HPLC (3-100% MeOH in 20 mM ammonium formate buffer pH 3.5 over 60 minutes). The collected fraction was lyophilized yielding a white solid as a formic salt (168.8 mg, 95%): ^{1}H NMR (400 MHz, D_2O) δ 8.41 (s, 1H, *H*COO^-), 8.31 (s, 1H, 2-*H*), 8.25 (s, 1H, 8-*H*), 6.06 (d, J = 5.3 Hz, 1H, 1'-*H*), 4.46 (dd, J = 5.3, 4.7 Hz, 1H, 2'-*H*), 4.44 – 4.38 (m, 1H, 3'-*H* under solvent peak), 3.95 – 3.84 (m, 4H, 5'-*H*, 5''-), 3.33 – 3.28 (m, 2H, $CH_2CH_2NH_3^+$); ^{13}C NMR (101 MHz, D_2O) δ 170.91, 154.63, 152.75, 148.40, 139.67, 119.11, 87.16, 83.37, 73.44, 70.67, 44.10, 39.27, 38.12; TOF MS ES (+) m/z $[M+H]^+$ calcd: 329.1129; found: 329.1136.

Synthesis of 2-((9-((2R,3R,4S,5S)-3,4-dihydroxy-5-(iodomethyl)tetrahydrofuran-2-yl)-9H-purin-6-yl)amino)ethan-1-aminium formate (5'-I,6-EDA-Ade,)



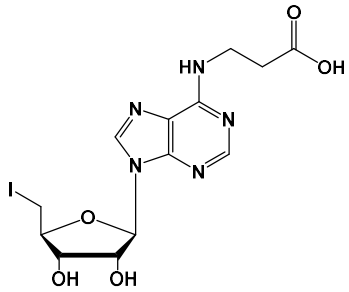
To a solution of 5'-I,6-Cl-Ade (150 mg, 0.38 mmol) in MeOH (1.5 mL), EDA (152 µL, 6 eq) was added and reaction was stirred for 2 hours. The MeOH was removed under reduced pressure and pH was adjusted to 3 with 1M HCl. Resulting solution was purified by preparative RP-HPLC (3-50% MeOH in 20 mM ammonium formate buffer pH 3.5 over 60 min). The collected fraction was lyophilized yielding an amorphic off-white solid as a formic salt (168.8 mg, 95%): ^{1}H NMR (400 MHz, DMSO-d_6) δ 8.43 – 8.38 (m, 3H, 2-*H*, 2x *H*COO$^-$), 8.26 (s, 1H, 8-*H*), 7.98 (br. s, 1H, N*H*CH$_2$CH$_2$NH$_3^+$), 5.93 (d, *J* = 5.8 Hz, 1H, 1'-*H*), 4.80 (dd, *J* = 5.4 Hz, 1H, 2'-*H*), 4.17 (dd, *J* = 5.1, 3.6 Hz, 1H, 3'-*H*), 4.02 – 3.95 (m, 1H, 4'-*H*), 3.75 – 3.57 (m, 3H, 5'-*H*, NHC*H*$_2$CH$_2$NH$_3^+$), 3.46 (dd, *J* = 10.4, 7.0 Hz, 1H, 5''-*H*), 2.97 (t, *J* = 6.3 Hz, 2H, NHCH$_2$C*H*$_2$NH$_3^+$); ^{13}C NMR (101 MHz, DMSO) δ 165.59, 152.56, 140.01, 129.38, 83.95, 79.20, 73.21, 72.88, 28.06, 7.97; TOF MS ES (+) m/z [M+H]$^+$ calcd: 421.0485; found: 421.0496.

Synthesis of 3-((9-((2R,3R,4S,5S)-5-(chloromethyl)-3,4-dihydroxytetrahydrofuran-2-yl)-9H-purin-6-yl)amino)propanoic acid (5'-Cl,6-β-Ala-Ade)



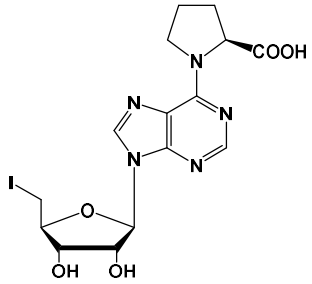
To a solution of 5',6-diCl-Ade (300 mg, 0.987 mmol) in MeOH (3 mL), β-alanine solution (167 mg, 1.9 eq.) in 800 µL of water and triethylamine (520 µL, 3.8 eq.) were added and reaction was stirred for 8 hours. pH was adjusted to 3 with 1M HCl. Resulting off-white precipitate was washed with cold MeOH and water, and used without further purification (329 mg, 93%): ^{1}H NMR (400 MHz, DMSO-d6) δ 12.22 (br. s, 1H, -COO*H*), 8.37 (s, 1H, 2-*H*), 8.27 (s, 1H, 8-*H*), 7.84 (s, 1H, N*H*CH_2), 5.95 (d, J = 5.6 Hz, 1H, 1'-*H*), 4.77 (dd, J = 5.6, 5.1 Hz, 1H, 2'-*H*), 4.25 (dd, J = 5.1, 3.5 Hz, 1H, 3'-*H*), 4.11 (ddd, J = 6.4, 5.1, 3.5 Hz, 1H, 4'-*H*), 3.96 (dd, J = 11.6, 5.1 Hz, 1H, 5'-*H*), 3.85 (dd, J = 11.6, 6.4 Hz, 1H, 5''-*H*), 3.78 – 3.61 (m, 2H, NHC$H_2$$CH_2$), 2.60 (t, J = 7.2 Hz, 2H, CH_2CH_2COOH); ^{13}C NMR (101 MHz, DMSO) δ 173.07, 154.43, 152.69 (8-C), 148.69, 139.77 (2-C), 119.63, 87.52 (1'-C), 83.72 (4'-C), 72.70 (2'-C), 71.29 (3'-C), 44.84 (5'-C), 36.07 (NH*C*$H_2$$CH_2$), 33.77 ($CH_2$*C*$H_2$COOH). TOF MS ES (-) m/z $[M-H]^-$ calcd: 356.0762; found: 356.0763.

Synthesis of 3-((9-((2R,3R,4S,5S)-5-(iodomethyl)-3,4-dihydroxytetrahydrofuran-2-yl)-9H-purin-6-yl)amino)propanoic acid (5'-I,6-β-Ala-Ade)



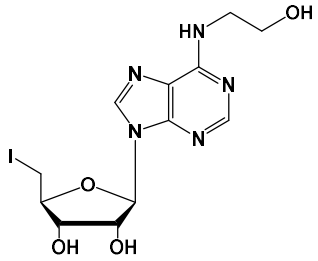
To a solution of 5',6-diCl-Ade (300 mg, 0.987 mmol) in MeOH (3 mL), β-alanine solution (167 mg, 1.9 eq.) in 800 µL of water and triethylamine (520 µL, 3.8 eq.) were added and reaction was stirred for 8 hours. pH was adjusted to 3 with 1M HCl. Resulting off-white precipitate was washed with cold MeOH and water, and used without further purification (329 mg, 93%); ^{1}H NMR (400 MHz, DMSO-d_6) δ 8.38 (s, 1H, 2-*H*), 8.25 (s, 1H, 8-*H*), 7.81 (br. s, 1H, N*H*CH_2CH_2COOH), 5.93 (d, *J* = 5.7 Hz, 1H, 1'-*H*), 4.81 (dd, *J* = 5.5 Hz, 1H, 2'-*H*), 4.18 (dd, *J* = 4.3 Hz, 1H, 3'-*H*), 4.03 – 3.95 (m, 1H, 4'-*H*), 3.68 (br. s, 2H, NHCH_2CH_2COOH), 3.61 (dd, *J* = 10.5, 5.9 Hz, 1H, 5'-*H*), 3.47 (dd, *J* = 10.5, 6.9 Hz, 1H, 5''-*H*), 2.59 (t, *J* = 7.2 Hz, 2H, $NHCH_2CH_2COOH$); ^{13}C NMR (101 MHz, DMSO) δ 139.89, 83.91, 73.18, 72.79, 54.93, 45.53, 8.98, 7.82. TOF MS ES (-) m/z $[M-H]^-$ calcd: 448.0118; found: 448.0128.

Synthesis of (9-((2R,3R,4S,5S)-3,4-dihydroxy-5-(iodomethyl)tetrahydrofuran-2-yl)-9H-purin-6-yl)-L-proline (5'-I,6-Pro-Ade)



To a solution of 5'-I,6-Cl-Ade (100 mg, 0.253 mmol) in MeOH (1 mL) and water (100 µL), L-Proline*HCl (174.5 mg, 6 eq.) and triethylamine (211 µL, 6 eq.) were added and reaction was stirred for 8 hours. Resulting solution was purified by preparative RP-HPLC (3– 100% MeOH in water over 60 min). The collected fraction was lyophilized yielding a white solid as (55 mg, 46 %): ^{1}H NMR (400 MHz, DMSO-d6) δ 12.55 (br. s, 1H, -COO*H*), 8.45 – 8.19 (m, 2H, 2-*H*, 8-*H*), 6.00 – 5.92 (m, 1H, 1'-*H*), 5.61 (d, J = 6.0 Hz, 1H, 2'-O*H*), 5.52 – 5.44 (m, 1H, 3'-O*H*), 5.37 – 5.29 (m, Hα-rot. 1), 4.85 – 4.76 (m, 1H, 2'-*H*), 4.69 – 4.61 (m, Hα-rot. 2), 4.25 – 4.12 (m, 2H, 3'-O*H*, Hδ), 4.05 – 3.95 (m, 1H, 4'-*H*), 3.84 – 3.67 (m, 1H, Hδ), 3.66 – 3.56 (m, 1H, 5'-*H*), 3.52 – 3.44 (m, 1H, 5''-*H*), 2.43 – 1.72 (m, 3H, Hβ, Hγ, one H under solvent peak). ^{13}C NMR (101 MHz, DMSO) δ 152.84, 152.50, 140.07, 139.83, 87.85, 84.35, 84.25, 73.65, 73.38, 73.14, 60.92, 60.06, 49.52, 47.94, 31.17, 29.18, 24.89, 22.42, 8.30; TOF MS ES (-) m/z $[M-H]^-$ calcd: 474.0274; found: 474.0270.

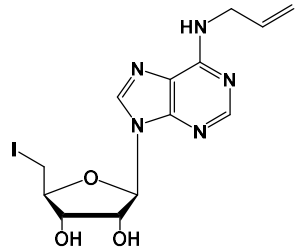
Synthesis of (2R,3R,4S,5S)-2-(6-((2-hydroxyethyl)amino)-9H-purin-9-yl)-5-(iodomethyl)tetrahydrofuran-3,4-diol (5'-I,6-ETA-Ade)



To a solution of 5'-I,6-Cl-Ade (100 mg, 0.253 mmol) in MeOH (1 mL), ethanolamine (91 µL, 6 eq.) and triethylamine (211 µL, 6 eq.) were added and reaction was stirred for 8 hours. Resulting solution was purified by semi-prep HPLC (3– 100% MeOH in

water over 60min). The collected fraction was lyophilized yielding a white solid as (80 mg, 75 %): ^{1}H NMR (500 MHz, DMSO-d6) δ 8.37 (s, 1H, 2-*H*), 8.23 (br. s, 1H, 8-*H*), 7.66 (br. s, 1H, N*H*CH$_2$), 5.93 (d, J = 5.7 Hz, 1H, 1'-*H*), 5.57 (d, J = 6.0 Hz, 1H, 2'-O*H*), 5.46 (d, J = 5.0 Hz, 1H, 3'-O*H*), 4.85 – 4.79 (m, 1H, 2'-*H*), 4.76 (br. s, 1H, CH$_2$O*H*), 4.21 – 4.15 (m, 1H, 3'-*H*), 4.02 – 3.96 (m, 1H, 4'-*H*), 3.65 – 3.50 (m, 5H, C*H*$_2$C*H*$_2$, 5'-*H*), 3.47 (dd, J = 10.4, 6.9 Hz, 1H, 5''-*H*); ^{13}C NMR (126 MHz, DMSO) δ 154.67, 152.62, 148.71, 139.81, 119.58, 87.53, 83.89, 73.18, 72.77, 59.68, 42.46, 7.83; TOF MS ES (+) m/z [M+H]$^{+}$ calcd: 422.0325; found: 422.0334

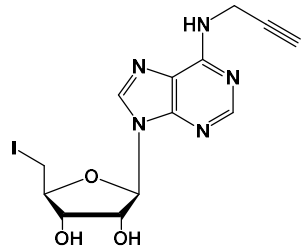
Synthesis of (2R,3R,4S,5S)-2-(6-(allylamino)-9H-purin-9-yl)-5-(iodomethyl)tetrahydrofuran-3,4-diol (5'-I,6-Allyl-Ade)



To a solution of 5'-I,6-Cl-Ade (100 mg, 0.1 mmol) in MeOH (1 mL), allylamine (114 µL, 6 eq.) and triethylamine (211 µL, 6 eq.) were added and reaction was stirred for 24 hours. The precipitate was then filtered off and washed with cold water, methanol and dried over P_2O_5 yielding a white solid (66 mg, 63%) which was used in the next step without further purification: ^{1}H NMR (400 MHz, DMSO-d_6) δ 8.39 (s, 1H, 2-*H*), 8.23 (br. s, 1H, 8-*H*), 8.04 (br. s, 1H, N*H*CH$_2$), 6.02 – 5.89 (m, 2H, 1'-*H*, CH$_2$C*H*=CH$_2$), 5.60 (d, *J* = 6.1 Hz, 1H, 2'-O*H*), 5.48 (d, *J* = 5.1 Hz, 1H, 3'-O*H*), 5.20 – 5.11 (m, 1H, CH$_2$CH=C*H*H), 5.08 – 5.02 (m, 1H, CH$_2$CH=C*H*H), 4.87 – 4.79 (m, 1H, 2'-*H*), 4.22 – 4.06 (m, 3H, 3'-*H*, NHC*H*$_2$CH=CH$_2$), 4.00 (ddd, *J* = 7.0, 5.9, 3.5 Hz, 1H, 4'-*H*), 3.62 (dd,

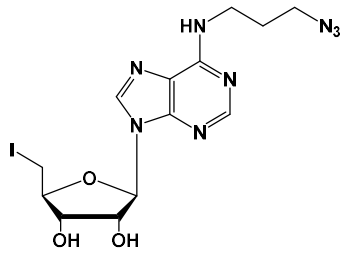
J = 10.4, 5.9 Hz, 1H, 5'-*H*), 3.48 (dd, *J* = 10.4, 7.0 Hz, 1H, 5''-*H*); ^{13}C NMR (101 MHz, DMSO) δ 154.89, 153.10, 140.29, 136.08, 119.96, 115.45, 87.95, 84.38, 73.64, 73.21, 42.40, 8.29; TOF MS ES (+) m/z $[M+H]^+$ calcd: 418.0376; found: 418.0386

Synthesis of (2S,3S,4R,5R)-2-(iodomethyl)-5-(6-(prop-2-yn-1-ylamino)-9H-purin-9-yl)tetrahydrofuran-3,4-diol (5'-I,6-Prop-Ade)



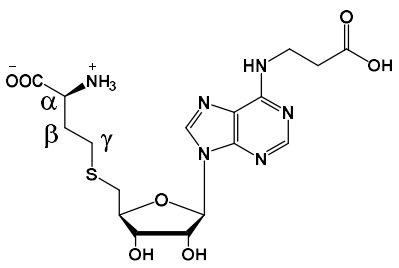
To a solution of 5'-I,6-Cl-Ade (40 mg, 0.1 mmol) in MeOH (2 mL), propargylamine (39 µL, 6 eq.) and triethylamine (84 µL, 6 eq.) were added and reaction was stirred for 48 hours. The precipitate was then filtered off and washed with cold water, methanol and dried over P_2O_5 yielding an off-white solid (28.4 mg, 67.7%) which was used in the next step without further purification: ^{1}H NMR (400 MHz, DMF-d_7) δ 8.46 (s, 1H, 2-*H*), 8.35 (s, 1H, 8-*H*), 8.13 (br. s, 1H, N*H*CH_2), 6.11 (d, *J* = 5.6 Hz, 1H, 1'-*H*), 5.79 (d, *J* = 5.7 Hz, 1H, 2'-O*H*), 5.56 (d, *J* = 4.9 Hz, 1H, 3'-O*H*), 5.04 (ddd, *J* = 5.7, 5.6, 5.2 Hz, 1H, 2'-*H*), 4.59 – 4.35 (m, 3H, 3'-*H*, NHCH_2C≡CH), 4.18 – 4.10 (m, 1H, 4'-*H*), 3.74 (dd, *J* = 10.4, 5.8 Hz, 1H, 5'-*H*), 3.60 (dd, *J* = 10.4, 6.8 Hz, 1H, 5''-*H*), 3.05 (t, *J* = 2.5 Hz, 1H, CH_2C≡C*H*); ^{13}C NMR (101 MHz, DMF) δ 152.95, 140.62, 88.78, 84.81, 82.04, 74.20, 73.82, 71.77, 7.15; MS ESI (+) $[M+H]^+$ calcd: 416.0214, found: 416.0231, $[M+Na]^+$ calcd: 438.0034, found: 438.0028

Synthesis of (2R,3R,4S,5S)-2-(6-((3-azidopropyl)amino)-9H-purin-9-yl)-5-(iodomethyl)tetrahydrofuran-3,4-diol (5'-I,6-PAA-Ade)



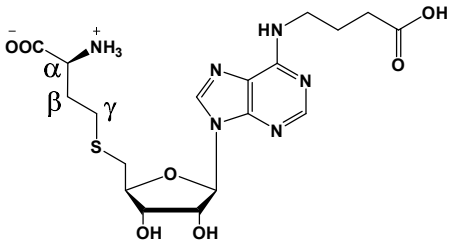
To a solution of 5'-I,6-Cl-Ade (40 mg, 0.1 mmol) in MeOH (0.4 mL), 3-azido-1-propanamine (60.7 mg, 6 eq.) and triethylamine (84µL, 6 eq.) were added and reaction was stirred for 8 hours. Resulting solution was diluted 10 fold with water/MeOH (1:1), pH adjusted to 7 with 1M HCl and purified by semi-prep HPLC (50– 100% MeOH in water over 60min). The collected fraction was lyophilized yielding a white solid (28.7 mg, 62 %): ^{1}H NMR (400 MHz, Methanol-d_4) δ 8.26 (m, J = 2.6 Hz, 2H, 2-*H*, 8-*H*), 6.01 (d, J = 5.2 Hz, 1H, 1'-*H*), 4.89 – 4.82 (m, 1H, 2'-*H* under solvent peak), 4.30 (dd, J = 5.4, 4.1 Hz, 1H, 3'-*H*), 4.06 (td, J = 5.7, 4.1 Hz, 1H, 4'-*H*), 3.68 (br. s, 2H, NHCH_2CH$_2$), 3.62 (dd, J = 10.7, 5.8 Hz, 1H, 5'-*H*), 3.50 (dd, J = 10.7, 5.8 Hz, 1H, 5''-*H*), 3.44 (t, J = 6.7 Hz, 2H, CH$_2$CH_2N$_3$), 1.94 (tt, J = 6.7 Hz, 2H, NHCH$_2$CH_2CH$_2$N$_3$); ^{13}C NMR (101 MHz, MeOD) δ 153.99, 141.00, 90.14, 85.19, 74.92, 74.87, 50.13, 6.26. ESI MS (+) m/z [M+H]$^+$ calcd: 461.0541, found: 461.0528

Synthesis of S-(((2S,3S,4R,5R)-5-(6-((2-carboxyethyl)amino)-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)-L-homocysteine (β-Ala-SAH)



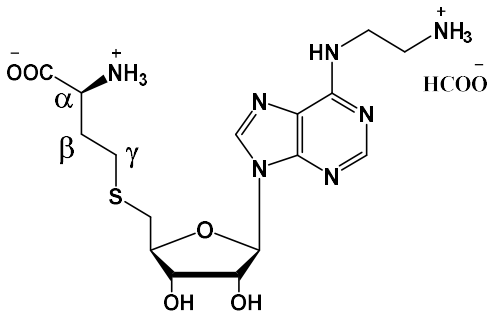
A solution of L-homocysteine (18.5 mg, 2 eq.) in 1M NaOH (200 µL, 3 eq.) was degassed under N_2 for 15 mins followed by addition of a solution of 5'-Cl,6- β-Ala-Ade (25 mg, 0.07 mmol) or 5'-I,6- β-Ala-Ade (30 mg, 0.067 mmol) in 300 µL of MeOH. The reaction was degassed for additional 10 mins and heated up to 100°C. The progress was monitored by HPLC. The reaction was quenched by addition of 1M HCl to pH 3-4. Reaction was purified by preparative RP-HPLC (3-100% MeOH in 20 mM ammonium formate buffer pH 3.5 over 60 min) and collected fraction was lyophilized to give β-Ala-SAH (12.1 mg, 38%, 24h) and (24.8 mg, 81%, 4.5h) respectively as a white solid. ^{1}H NMR (400 MHz, DMSO-d6) δ 8.37 (s, 1H, 2-*H*), 8.25 (s, 1H, 8-*H*), 7.80 (s, 1H, N*H*CH_2), 5.90 (d, J = 5.6 Hz, 1H, 1'-*H*), 4.79 – 4.67 (m, 1H, 2'-*H*), 4.20 – 4.12 (m, 1H, 3'-*H*), 4.09 – 3.97 (m, 1H, 4'-*H*), 3.68 (br. s, 2H, NHC*H*$_2$$CH_2$), 3.35 – 3.29 (m, 1H, Hα), 2.91 (dd, J = 13.8, 6.0 Hz, 1H, 5'-*H*), 2.80 (dd, J = 13.8, 6.9 Hz, 1H, 5''-*H*), 2.66 – 2.54 (m, 4H, 2Hγ, NHCH$_2$C*H*$_2$COOH), 2.05 – 1.91 (m, 1H, Hβ), 1.88 – 1.73 (m, 1H, Hβ); ^{13}C NMR (101 MHz, DMSO) δ 173.15, 170.14, 152.64, 139.81, 87.46, 83.68, 72.80, 72.65, 52.95, 33.85, 31.34, 28.10; ESI MS (+) m/z $[M+H]^+$ calcd: 457.1500, found: 457.1519

Synthesis of S-(((2S,3S,4R,5R)-5-(6-((3-carboxypropyl)amino)-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)-L-homocysteine (GABA-AdoHcy)



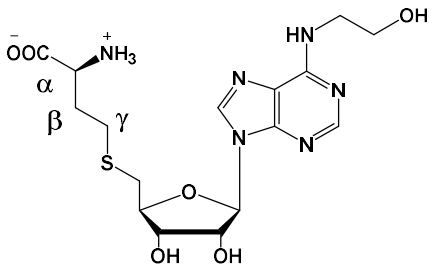
A solution of L-homocysteine (18 mg, 2 eq.) in 1M NaOH (190 µL, 3 eq.) was degassed under N_2 for 15 mins followed by addition of a solution of 5'-Cl,6-GABA-Ade (25 mg, 0.067 mmol) or 5'-I,6-GABA-Ade (30 mg, 0.065 mmol) in 310 µL of MeOH. The reaction was degassed for additional 10 mins and heated up to 100°C. The progress was monitored by HPLC. The reaction was quenched by addition of 1M HCl to pH 3-4. Reaction was purified by preparative RP-HPLC (3-100% MeOH in 20 mM ammonium formate buffer pH 3.5 over 60 min) and collected fraction was lyophilized to give GABA-AdoHcy (11 mg, 35%, 24h) and (22.6 mg, 74%, 4.5h) respectively as a white solid; ^{1}H NMR (400 MHz, DMSO-d6) δ 8.36 (s, 1H, 2-*H*), 8.22 (s, 1H, 8-*H*), 7.94 (br. s, 1H, N*H*CH_2), 5.89 (d, J = 5.6 Hz, 1H, 1'-*H*), 4.78 – 4.67 (m, 1H, 2'-*H*), 4.18 – 4.12 (m, 1H, 3;-*H*), 4.06 – 3.98 (m, 1H, 4'-*H*), 3.49 (br. s, 2H, NHCH_2CH_2), 3.36 – 3.28 (m, 1H, Hα), 2.91 (dd, J = 13.8, 5.9 Hz, 1H, 5'-*H*), 2.81 (dd, J = 13.8, 6.9 Hz, 1H, 5''-*H*), 2.63 (t, J = 7.7 Hz, 2H, Hγ), 2.26 (t, J = 7.4 Hz, 2H, $NHCH_2CH_2CH_2$COOH), 2.05 – 1.92 (m, 1H, Hβ), 1.88 – 1.75 (m, 3H, Hβ, $NHCH_2CH_2CH_2$COOH); ^{13}C NMR (101 MHz, DMSO) δ 174.61, 170.28, 152.68, 139.58, 87.45, 83.75, 72.85, 72.63, 52.97, 33.85, 31.69, 31.40, 28.16; MS ESI $[M+H]^+$ calcd: 471.1662, found: 471.1666

Synthesis of (S)-2-ammonio-4-((((2S,3S,4R,5R)-5-(6-((2-ammonioethyl)amino)-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)thio)butanoate formate (EDA-AdoHcy)



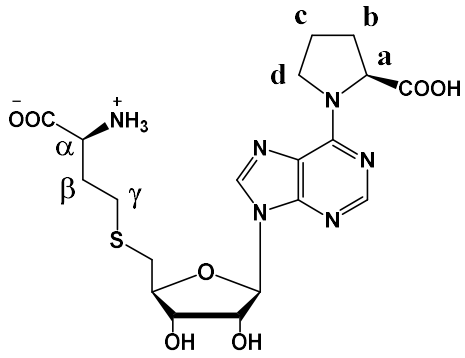
A solution of L-homocysteine (36.2 mg, 2 eq.) in 1M NaOH (403 µL, 3 eq.) was degassed under N_2 for 15 mins followed by addition of a solution of 5'-Cl,6-EDA-Ade (50 mg, 0.134 mmol) or 5'-I,6-EDA-Ade (62.5 mg, 0.134 mmol). The reaction was degassed for additional 10 mins and heated up to 100°C. The progress was monitored by HPLC. The reaction was quenched by addition of 1M HCl to pH 3-4. Reaction was purified by preparative RP-HPLC (3-50% MeOH in 20 mM ammonium formate buffer pH 3.5 over 60 min) and collected fraction was lyophilized to give EDA-AdoHcy (23.5 mg, 37%, 5h) and (42.5 mg, 67%, 10min) respectively as a white solid. ^{1}H NMR (400 MHz, DMSO-d_6) δ 8.43 – 8.39 (m, 3H, 2-*H*, 2x*H*COO$^-$), 8.26 (s, 1H, 8-*H*), 8.05 (s, 1H, N*H*CH$_2$), 5.90 (d, *J* = 5.5 Hz, 1H, 1'-*H*), 4.71 (dd, *J* = 5.5, 5.0 Hz, 1H, 2'-*H*), 4.15 (dd, *J* = 5.0, 3.9 Hz, 1H, 3'-*H*), 4.03 (ddd, *J* = 6.9, 5.7, 3.9 Hz, 1H, 4'-*H*), 3.68 (br. s, *J* = 8.5 Hz, 2H, NHC*H*$_2$CH$_2$NH$_3^+$), 3.32 (dd, *J* = 6.9, 5.3 Hz, 1H, Hα), 3.00 (t, *J* = 6.2 Hz, 2H, NHCH$_2$C*H*$_2$NH$_3^+$), 2.90 (dd, *J* = 13.9, 5.7 Hz, 1H, 5'-*H*), 2.81 (dd, *J* = 13.9, 6.9 Hz, 1H, 5''-*H*), 2.62 (t, *J* = 7.7 Hz, 2H, Hγ), 2.03 – 1.89 (m, 1H, Hβ), 1.88 – 1.76 (m, 1H, Hβ); ^{13}C NMR (101 MHz, DMSO) δ 165.37, 152.29, 139.65, 87.23, 83.67, 72.66, 72.35, 52.66, 39.11, 38.78, 33.60, 31.21, 27.94. TOF MS ES (+) m/z [M+H]$^+$ calcd: 428.1716, found: 428.1724.

Synthesis of (S)-2-ammonio-4-((((2S,3S,4R,5R)-3,4-dihydroxy-5-(6-((2-hydroxyethyl)amino)-9H-purin-9-yl)tetrahydrofuran-2-yl)methyl)thio)butanoate (EtA-AdoHcy)



A solution of L-homocysteine (25.7 mg, 2 eq.) in 1M NaOH (210 µL, 2.2 eq.) was degassed under N_2 for 15 mins followed by addition of a solution 5'-I,6-ETA-Ade (40 mg, 0.095 mmol) in 100 µL MeOH. The reaction was degassed for additional 10 mins and heated up to 100°C. The progress was monitored by HPLC. The reaction was quenched by addition of 1M HCl to pH 3-4. Reaction was purified by preparative RP-HPLC (3-100% MeOH in 20 mM ammonium formate buffer pH 3.5 over 60 min) and collected fraction was lyophilized to give (16.8 mg, 41.3%,4 hours) as a white solid. ^{1}H NMR (400 MHz, DMSO-d_6) δ 8.38 (s, 1H, 2-*H*), 8.23 (s, 1H, 8-*H*), 7.75 (br. s, 1H, N*H*CH_2), 5.90 (d, *J* = 5.7 Hz, 1H, 1'-*H*), 4.75 – 4.66 (m, 1H, 2'-*H*), 4.19 – 4.10 (m, 1H, 3'-*H*), 4.09 – 3.99 (m, 1H, 4'-*H*), 3.63 – 3.49 (m, 4H, NHCH_2CH_2OH), 3.47 – 3.11 (m, 1H, Hα), 2.91 (dd, *J* = 13.8, 6.0 Hz, 1H, 5'-*H*), 2.81 (dd, *J* = 13.8, 7.0 Hz, 1H, 5''-*H*), 2.69 – 2.58 (m, 2H, Hγ), 2.08 – 1.93 (m, 1H, Hβ), 1.92 – 1.77 (m, 1H, Hβ); ^{13}C NMR (101 MHz, DMSO) δ 170.36, 154.64, 152.68, 148.71, 139.67, 119.50, 87.41, 83.72, 72.91, 72.64, 59.68, 53.02, 42.66, 33.91, 31.42, 28.12; TOF MS ES (+) m/z $[M+H]^+$ calcd: 429.1556, found: 429.1564.

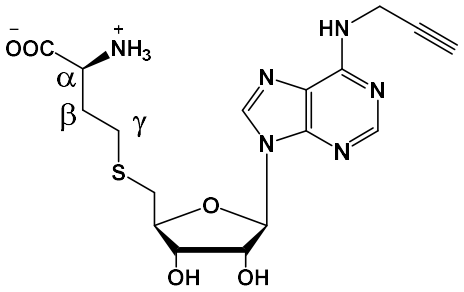
Synthesis of (S)-2-ammonio-4-((((2S,3S,4R,5R)-5-(6-((S)-2-carboxypyrrolidin-1-yl)-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)thio)butanoate (Pro-AdoHcy)



A solution of L-homocysteine (17.1 mg, 2 eq.) in 1M NaOH (190 µL, 2 eq.) was degassed under N_2 for 30 mins followed by addition of 5'-I,6-Pro-Ade (30 mg, 0.063 mmol.) in 110 µL of MeOH. The reaction was degassed for additional 10 mins and heated up to 100°C. The progress was monitored by HPLC. The reaction was quenched by addition of 1M HCl to pH 3-4. Reaction was purified by preparative RP-HPLC (3-100% MeOH in 20 mM ammonium formate buffer pH 3.5 over 60 min) and collected fraction was lyophilized yielding to give (29.9 mg, 98.2 %, 3 hours) as a white solid; ^{1}H NMR (400 MHz, D_2O+0.1% TFA) δ 8.32 – 8.07 (m, 2H, 2-*H*, 8-*H*), 6.04 (d, *J* = 4.8 Hz, 1H, 1'-*H*), 5.16 (dd, *J* = 8.9, 2.9 Hz, 1H, Ha rot. 1), 4.79 (s, 3H, 2'-*H*, two proton under solvent peak), 4.56 (dd, *J* = 8.3, 3.4 Hz, 1H, Ha rot. 2), 4.44 – 4.34 (m, 1H, 3'-*H*), 4.34 – 4.26 (m, 1H, 4'-*H*), 4.25 – 4.00 (m, 1H), 3.87 – 3.76 (m, 1H), 3.74 – 3.61 (m, 1H), 3.08 – 2.88 (m, 2H, 5'-*H,* 5''-*H*), 2.72 – 2.57 (m, 2H), 2.48 – 1.85 (m, 4H); ^{13}C NMR (101 MHz, D_2O+0.1% TFA) δ 179.48, 174.01, 170.41, 151.45, 150.85, 149.86, 148.54, 139.20, 138.77, 119.86, 87.44, 83.34, 73.45, 73.25, 72.26, 64.27,

63.10, 53.67, 50.08, 48.90, 33.37, 31.29, 30.30, 29.65, 27.82, 24.33, 22.39. TOF MS ES (+) m/z $[M+H]^+$ calcd: 483.1662, found: 483.1675

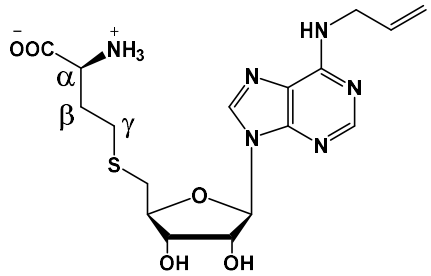
Synthesis of (S)-2-ammonio-4-((((2S,3S,4R,5R)-3,4-dihydroxy-5-(6-(prop-2-yn-1-ylamino)-9H-purin-9-yl)tetrahydrofuran-2-yl)methyl)thio)butanoate (Prop-AdoHcy)



A solution of L-homocysteine (7 mg, 2 eq.) in 2M NaOH (37.2 µL, 2.9 eq.) was degassed under N_2 for 15 mins followed by addition of a solution 5'-I,6-Prop-Ade (11 mg, 0.026 mmol) in 30 µL DMF. The reaction was degassed for additional 10 mins and heated up to 100°C. The progress was monitored by HPLC. The reaction was quenched by addition of 1M HCl to pH 3-4. Reaction was purified by preparative RP-HPLC (3-100% MeOH in 20 mM ammonium formate buffer pH 3.5 over 60 min) and collected fraction was lyophilized to give (7 mg, 64%,1h 30 min) as a white solid. ^{1}H NMR (400 MHz, DMSO-d_6) δ 8.42 (s, 1H, 2-*H*), 8.30 (s, 1H, 8-*H*), 8.23 (br. s, 1H, N*H*CH_2C≡CH), 5.91 (d, *J* = 5.8 Hz, 1H, 1'-*H*), 4.74 (dd, *J* = 5.4 Hz, 1H, 2'-*H*), 4.25 (br. s, 2H, NHC*H*$_2$C≡CH), 4.17 – 4.13 (m, 1H, 3'-*H*), 4.06 – 3.99 (m, 1H, 4'-*H*), 3.29 (dd, *J* = 7.1, 5.2 Hz, 1H, Hα), 3.03 (t, *J* = 2.4 Hz, 1H, NHCH$_2$C≡C*H*), 2.92 (dd, *J* = 13.7, 6.2 Hz, 1H, 5'-*H*), 2.80 (dd, *J* = 13.7, 6.9 Hz, 1H, 5''-*H*), 2.63 (t, *J* = 7.7 Hz, 2H, Hγ), 2.06 – 1.76 (m, 2H, Hβ); ^{13}C NMR (101 MHz, DMSO) δ 169.96, 164.88, 152.51, 140.20,

87.44, 83.65, 81.90, 72.80, 72.62, 53.02, 33.86, 31.41, 28.12; TOF MS ES (+) m/z $[M+H]^+$ calcd: 445.1270, found: 445.1277.

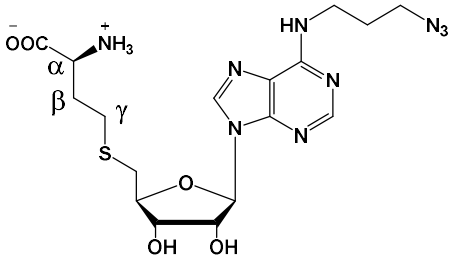
Synthesis of (S)-4-((((2S,3S,4R,5R)-5-(6-(allylamino)-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)thio)-2-ammoniobutanoate (Allyl-AdoHcy)



A solution of L-homocysteine (15.8 mg, 2 eq.) in 2M NaOH (84.6 µL, 2.9 eq.) was degassed under N_2 for 15 mins followed by addition of a solution 5'-I,6-Allyl-Ade (25 mg, 0.060 mmol) in 60 µL DMF. The reaction was degassed for additional 10 mins and heated up to 100°C. The progress was monitored by HPLC. The reaction was quenched by addition of 1M HCl to pH 3-4. Reaction was purified by preparative RP-HPLC (3-100% MeOH in 20 mM ammonium formate buffer pH 3.5 over 60 min) and collected fraction was lyophilized to give (24.7 mg, 97%,1h 30 min) as a white solid. ^{1}H NMR (400 MHz, DMSO-d_6) δ 8.38 (s, 1H, 2-*H*), 8.22 (s, 1H, 8-*H*), 8.03 (br. s, 1H, N*H*$CH_2CH{=}CH_2$), 6.01 – 5.87 (m, 2H, 1'-*H*, $NHCH_2CH{=}CH_2$), 5.19 – 5.10 (m, 1H, $NHCH_2CH{=}CHH$), 5.08 – 5.02 (m, 1H, $NHCH_2CH{=}CHH$), 4.74 (dd, *J* = 5.5 Hz, 1H, 2'-*H*), 4.19 – 4.06 (m, 3H, 3'-*H*, $NHCH_2CH{=}CH_2$), 4.05 – 3.97 (m, 1H, 4'-*H*), 3.31 (dd, *J* = 7.1, 5.2 Hz, 1H, Hα), 2.92 (dd, *J* = 13.7, 6.2 Hz, 1H, 5'-*H*), 2.80 (dd, *J* = 13.7, 6.9 Hz, 1H, 5''-*H*), 2.63 (t, *J* = 7.8 Hz, 2H, Hγ), 2.06 – 1.76 (m, 2H, 2xHβ); ^{13}C NMR (101 MHz, DMSO) δ 170.03, 163.92, 152.64, 139.77, 135.70, 115.00, 87.39, 83.60, 72.76, 72.64,

52.99, 33.87, 31.37, 28.09; TOF MS ES (+) m/z $[M+H]^+$ calcd: 425.1607, found: 425.1618.

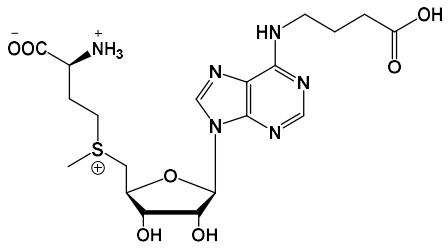
Synthesis of (S)-2-ammonio-4-((((2S,3S,4R,5R)-5-(6-((3-azidopropyl)amino)-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)thio)butanoate (PAA-AdoHcy)



A solution of L-homocysteine (58.8 mg, 4 eq.) in 1M NaOH (430 µL, 4 eq.) was degassed under N_2 for 30 mins followed by addition of 5'-I,6-PAA-Ade (50 mg, 0.108 mmol) in 500 µL of MeOH. The reaction was degassed for additional 10 mins and heated up to 100°C. The progress was monitored by HPLC. The reaction was quenched by addition of 1M HCl to pH 3-4. Reaction was purified by preparative RP-HPLC (3-100% MeOH in 20 mM ammonium formate buffer pH 3.5 over 60 min) and collected fraction was lyophilized yielding PAA-AdoHcy (32.9 mg, 65 %, 5 hours) as a white solid; ^{1}H NMR (400 MHz, DMSO-d_6) δ 8.37 (s, 1H, 2-*H*), 8.23 (s, 1H, 8-*H*), 7.96 (br. s, 1H, N*H*CH_2), 5.89 (d, *J* = 5.8 Hz, 1H, 1'-*H*), 4.72 (dd, *J* = 5.8, 5.1 Hz, 1H, 2'-*H*), 4.14 (dd, *J* = 5.1, 3.6 Hz, 1H, 3'-*H*), 4.05 – 3.99 (m, 1H, 4'-*H*), 3.54 (br. s, 2H, NHC$H_2$$CH_2$), 3.42 (t, *J* = 6.7 Hz, 2H, $NHCH_2CH_2CH_2N_3$), 3.30 (dd, *J* = 7.1, 5.2 Hz, 1H, Hα), 2.91 (dd, *J* = 13.8, 6.2 Hz, 1H, 5'-*H*), 2.79 (dd, *J* = 13.8, 6.9 Hz, 1H, 5''-*H*), 2.62 (t, *J* = 7.8 Hz, 2H, Hγ), 2.05 – 1.93 (m, 1H, Hβ), 1.89 – 1.78 (m, 3H, Hβ, $NHCH_2CH_2CH_2N_3$); ^{13}C NMR (101 MHz, DMSO) δ 170.20, 165.64, 154.62, 152.71,

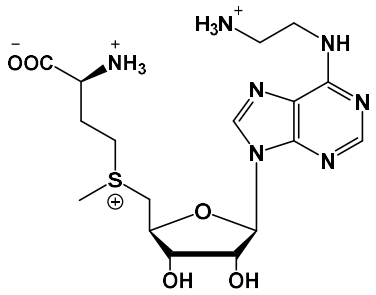
139.73, 87.42, 83.66, 72.85, 72.66, 53.09, 48.59, 37.21, 33.92, 31.46, 28.44, 28.13; TOF MS ES (+) m/z $[M+H]^+$ calcd: 468.1778, found: 468.1782

Synthesis of (2S)-2-ammonio-4-((((2S,3S,4R,5R)-5-(6-((3-carboxypropyl)amino)-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)(methyl)sulfonio)butanoate (GABA-AdoMet)



To a solution of GABA-AdoHcy (10 mg, 0.021 mmol) in $HCOOH/CH_3COOH$ (1:1, 200 µL), iodomethane (13 µL, 10 eq) was added slowly at 0°C. Then, a solution of $AgClO_4$ (4.3 mg, 1 eq.) in 200 µL of $HCOOH/CH_3COOH$ (1:1) mixture was added. The reaction mixture was stirred for 2 hours at 30°C in dark. Additional 13 µL of iodomethane and 1.3 mg of $AgClO_4$ was added and stirred for another 2 hours. The resulting mixture was diluted 10 times with water, AgI precipitate was was filtered off using 0.45 µm membrane filter and CH_3I was extracted with diethyl ether. Then, traces of organic solvent were removed under reduced pressure. The solution was purified by preparative RP-HPLC (3-100% MeOH in 20 mM ammonium formate buffer pH 3.5 over 60 min) and the collected fraction was lyophilized. The product (0.0145 mmol, 69%) was stored at -20°C in 0.1% formic acid solution. TOF MS ES (+) m/z M calcd: 485.1819, found: 485.1832.

Synthesis of (2S)-ammonio-4-((((2S,3S,4R,5R)-5-(6-((2-ammonioethyl)amino)-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)(methyl)sulfonio)butanoate (EDA-AdoMet)



To a solution of EDA-AdoHcy (6 mg, 0.0127 mmol) in HCOOH/CH_3COOH (1:1, 80 µL), iodomethane (12.5 µL, 20 eq) was added slowly at 0°C. Then, a solution of $AgClO_4$ (2.2 mg, 1 eq.) in 80 µL of HCOOH/CH_3COOH (1:1) mixture was added. The reaction mixture was stirred for 2 hours at 30°C in dark. The resulting mixture was diluted 10 times with water, AgI precipitate was filtered off using 0.45 µm membrane filter and CH_3I was extracted with diethyl ether. Then, traces of organic solvent were removed under reduced pressure. The solution was purified by preparative RP-HPLC (3-30% MeOH in 20 mM ammonium formate buffer pH 3.5 over 30 min then 30-100% over 30 min) and the collected fraction was lyophilized. The product (0.0044 mmol, 35%) was stored at -20°C in 0.1% formic acid solution. TOF MS ES (+) m/z M^+ calcd: 442.1873, found: 442.1877.

Restriction assay using M.TaqI, pUC19 and synthetic cofactors (EDA-AdoMet, GABA-AdoMet)

For pUC19 restriction assay, a master mix was prepared on ice of 50 ng/μL pUC19, 250 μM **cofactor analogue (GABA-AdoMet or EDA-AdoMet)** in NEB Cutsmart buffer pH = 8.5. The master mix was split into 6 vials (20 μL first and 10 μL other five). To the first vial, M.TaqI was added to reach the concentration of 37.5 ng/μL. Then a half of the sample one was transferred into sample 2. It was continued until the last sample. The controls were prepared with the same concentrations except the corresponding components were omitted. Two additional controls were set up with 80 μM AdoMet and the same concentrations of other components. All the samples were incubated for 1 hour at 50°C. 0.5 μL r.TaqI was added into all samples except one of the AdoMet controls and then samples were incubated for 1 hour at 50°C. Next, 0.5 μL of Proteinase K (20 mg/mL) was added to each sample and all were incubated for 1 hour at 50°C. Samples were ran on 1% agarose gel (120V for 40 minutes).

II.2. Alkylation of unmethylated DNA in fixed cells with extrinsic methyltransferases

II.2.1. Introduction

II.2.1.1. Current approaches for the analysis of epigenetic modifications and chromatin structure

Epigenetic modifications of chromatin as well as its conformation and accessibility are studied using multiple approaches, of which the most popular are microscopy-based and sequencing-based techniques[223-226]. These usually allow a bulk analysis and provide the information on the distribution and interaction of epigenetic modifications alongside chromatin organisation in a population of cells. Some of the techniques can be applied at a single cell level which allows studies of intercellular interactions and a genomic heterogeneity, which is important in many diseases including cancer, and the study of epigenetic alterations in the early development of embryos.[223, 227]

The sequencing-based techniques for epigenetics analysis use a range of approaches to modify the biological sample prior to sequencing (Fig. 48).

The gold standard for genome-wide analysis of m5C is bisulphite sequencing (BS-seq). In this technique, purified genomic DNA is treated with bisulphite salts which converts unmethylated cytosine residues to uracil, leaving 5-methylcytosine unaffected (Fig. 48 A).[228] The modified DNA is amplified using the polymerase chain reaction (PCR) which replaces all U → T and m5C → C in the newly synthesized DNA strands. By comparing two sequencing experiments, m5C in the original DNA template sequence can be identified. However, a significant drawback of the bisulfite treatment

is that the chemical reaction degrades DNA, and hence large quantities of material are typically required for robust epigenomic profiling.

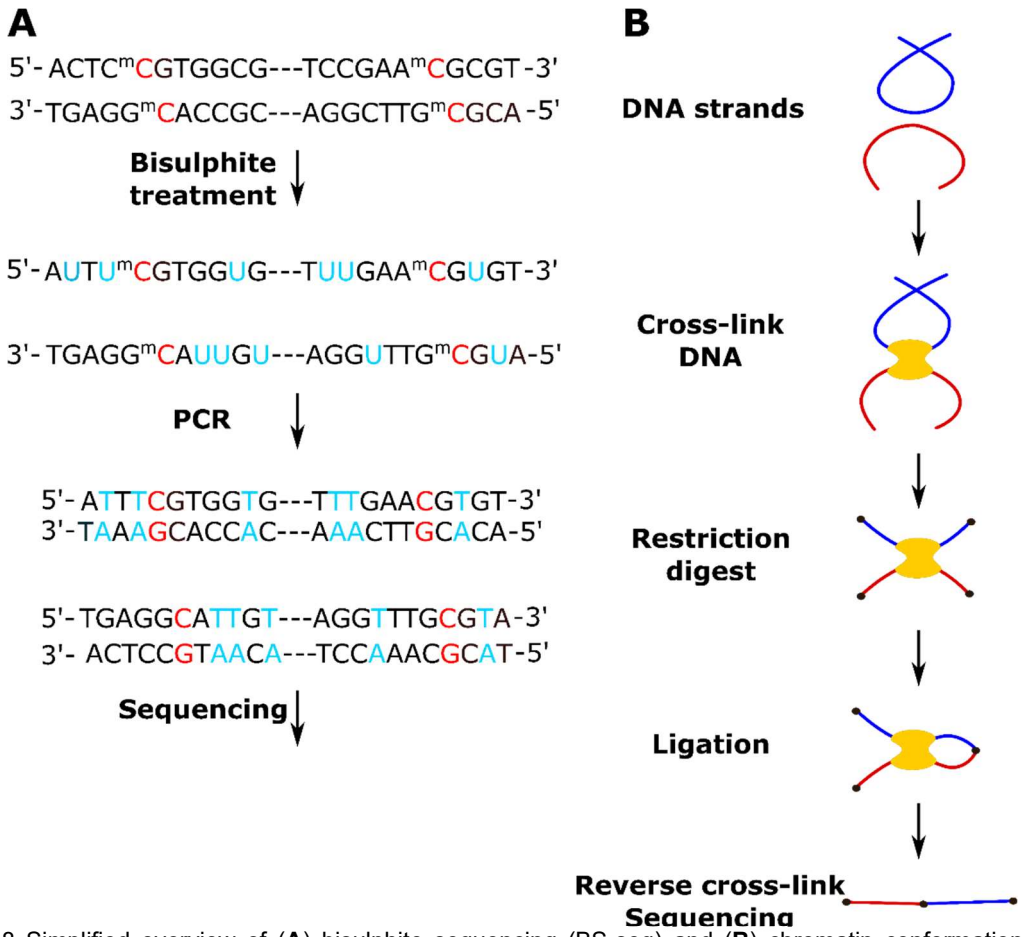


Fig. 48 Simplified overview of (**A**) bisulphite sequencing (BS-seq) and (**B**) chromatin conformation capture techniques.

Another approach that is widely used in the analysis of m5C, 5hmC and DNA interaction with histone modifications is immunoprecipitation. This approach applies specific antibodies that recognise 5mC or 5hmC on fragmented genomic DNA molecules (MeDIP-seq, hMeDIP-seq) when the antibody binds to these targets, they can then be precipitated from solution, separating the fragments of interest from the rest of the genome. The isolated material can then be sequenced.[229, 230] However, these approaches have some limitations. MeDIP-seq does not differentiate CpG and non-CpG modifications and is typically biased towards hypermethylated regions of

DNA.[231] Some antibodies also display nonspecific interaction which with the DNA, which complicates the analysis of data.

Many sequencing based methods, such as DNase-seq, ATAC-seq and FAIRE-seq, use enzymatic restriction of DNA to study its accessibility.[232, 233] These methods have similar principle in which an enzyme is used to fragment chromatin in a nucleus where the sites are then marked to be recognised during sequencing. It is assumed that the regions of the genome that are accessible to the enzymes are in the form of open chromatin. Though, these techniques can be used to study the epigenome in both a genome-wide context and specific regions with single base pair resolution, they lack information on the chromatin conformation and spatial interactions of distant regions of the genome.

Chromatin conformation capture (3C, 4C, 5C, Hi-C) techniques that combine chemical crosslinking, digestion and sequencing were developed to study the three-dimensional organisation of chromatin (Fig. 48 B).[234, 235] In these methods, spatially adjacent fragments of chromatin are chemically cross-linked, e.g. with formaldehyde, followed by shearing and re-ligation in which new connections between fragments in a close proximity are formed. Subsequently, the crosslinking is removed and DNA is sheared prior to sequencing. Contact matrices are produced as a result which map intra- and interchromosomal interactions in the nucleus. These methods led to identification of topologically associated domains (TADs) and organisation of the chromatin in compartments A and B (active and inactive chromatin).[96, 235] The Hi-C method was also applied in a single-cell analysis in which a three-dimensional structure of chromosome X was reconstructed[236] and later a genome-wide three-dimensional interactions were analysed in single cells.[237] Recently, Hi-C was combined with BS-

seq (Methyl-HiC) to map not only the conformation of chromatin but to also get an insight into methylation patterns.[238]

Chromatin conformation capture techniques enables a detailed analysis of the long-distance interactions of DNA in a population of cells at the single-cell level. However, the preparation of sample is a complex multistep process, especially when single cell precision is needed. For example, these experiments require the isolation of nuclei prior to other treatments. It is currently unclear that the internal structure of the nucleus in unaffected during this and subsequent steps of the process. Moreover, the analysis and interpretation of the obtained data is challenging not only due to the large datasets acquired during the experiment (up to 7.3 billion reads per sample) but also the complexity of data which contain information on compartments, TADs and various other nucleolar interactions.[239]

Epigenetic feature	Method	Approach	Single cell analysis
m5C	BS-seq	Chemical modification	Yes
	MeDIP-seq	Immunoprecipitation	No
	Methyl-seq	Restriction enzyme	Yes
5hmC	oxBS-seq	Chemical modification	No
	TAB-seq	Enzymatic modification	Yes
	hMeDIP-seq	Immunoprecipitation	No
	Aba-seq	Restriction enzyme	Yes

Histone modification	ChIP-seq	Immunoprecipitation	Yes
Transcription factor binding	DamID	Cell transfection and specific modifications of chromatin (markers)	Yes for nuclear lamina interactions
DNA accessibility	DNase-seq ATAC-seq	Enzymatic DNA fragmentation	Yes
	FAIRE-seq	Chemical crosslinking, mechanical fragmentation	No
Nucleosome positioning	MNase-seq	Enzymatic digestion	Yes
	NOMe-seq	Enzymatic methylation (markers) then BS-seq	Yes
Three-dimensional organisation	HiC	Chemical crosslinking, enzymatic digestion	Yes
m5C and three-dimensional organisation	Methyl-HiC	Combination of Hi-C and BS-seq	Yes

Table 2 Overview of epigenetics techniques. m5C – 5-methylcytosine, 5hmC – 5-hydroxymethylcytosine

For decades, microscopy-based methods played an important role in the study of epigenetic modifications and chromatin structure including a first indication of the presence of nucleosomes in the structure of chromatin shown in 1974 by A. Olins and D.Olins.[240] Sequencing-based methods have recently been more and more popular and replaced microscopic methods in many cases as they can provide a high resolution information on multiple features at the same time.[226] Nevertheless,

the simplicity and quick sample preparation and data acquisition of microscopy-based methods, as well as availability for bulk and a single cell analysis are reasons that often make them a first choice in the epigenetic studies.

The most common and widely used methods are immunofluorescence (IF) and fluorescent in situ hybridization (FISH). These can be applied to visualise the organization, global and local distribution of chromatin, gene expression levels and (co)localization of epigenetic modifications in fixed cells.[241] IF utilizes antibodies that bind to a target modification across cell, which then can be visualized by fluorescently labelled secondary antibodies that recognise the primary antibody. FISH is based on the affinities of complementary DNA and RNA oligonucleotides which can be designed to target specific loci in cells. Variants of these methods were developed to study epigenetics modifications, e.g. to visualize DNA methylation patterns at specific satellite repeats[242] or hypermethylation of cytosine residues in a tumour suppressor gene in cancer cells.[243]

One of the drawbacks of sequencing-based methods is their inability to capture the dynamics of the epigenetic regulation, as cells or tissues need to be fixed or lysed prior the analysis. In recent years, methods to image chromatin organisation,[244] dynamic changes of histones modifications,[245] in a regulation of transcription by changes in a level of H3K27ac[246], histone modifications and the methylation of cytosine in live cells have been reported.[247]

One of the methods to study gene expression in live cells is MS2-MCP method.[248, 249] A gene of interest is modified in such a way that upon expression, mRNA with the known number of bacteriophage MS2 stem-loops is synthesized. MS2 stem-loops are

specifically recognised and stoichiometrically bound by co-expressed fluorescently tagged MS2 coat protein (MCP). This enables the quantification a newly synthesized mRNA, and hence determine the expression level of the gene of interest. MS2-MCP method was also recently combined with CRISPR-dCas9 system to visualise specific genomic loci in live cells.[250] In this method CRISPR-dCas9 system is responsible for localisation of the specific genomic loci with single-guide RNAs, tagged with MS2 stem-loops, which then are recognised and bound by fluorescently labelled MCP.

However, these methods require a specific probe which usually needs to be designed for each imaged feature. Such probes consist of fluorescently labelled proteins that are expressed in cells and bind specifically to a target motif. In addition, a background signal from cellular structures, non-specific interactions of the probe in the cell or a weak signal due to a low concentration of the targeted biomolecule can be problematic.

Although a great repertoire of the methods to study epigenetic modification and chromatin organisation exists, they are either limited to one type of information provided, e.g. methylation status or chromatin structure, or require a complex multistep processing. A recent development of Methyl-Hi-C shows that there is a need for methods that can provide information for a multi-level analysis. Moreover, many techniques are based on the specificity and strength of the affinity of antibodies to their target which might be compromised during complex post-labelling processing. Hence, methods of selective introduction of covalently bound tags may be advantageous in many cases.

II.2.2. Aim and method overview

The methylation of cytosine is the most abundant methylation mark in mammalian DNA. It occurs predominantly within the CpG dinucleotide sequence which is of great interest due to its role in many biological processes e.g. gene imprinting or regulation of gene expression. This epigenetic mark on DNA changes the activity of many proteins such as transcription factors or chromatin modellers. The epigenetic dysregulation and subsequent changes in chromatin organisation are key factors in many diseases including cancer (See I.1.3.5). Therefore, we aimed to design a method of studying DNA methylation in the context of chromatin architecture in cancer cells. We decided to apply methyltransferase directed DNA alkylation using synthetic AdoMet analogues to visualise the organisation of unmethylated and accessible genome in the cell nucleus. In the first step, the unmethylated and accessible genomic DNA, in fixed and permeabilized cells, is alkylated with a methyltransferase and a synthetic cofactor of choice bearing a clickable linker. Subsequently, the tagged DNA

is fluorescently labelled via a click reaction and cells imaged under the microscope (Fig. 49).

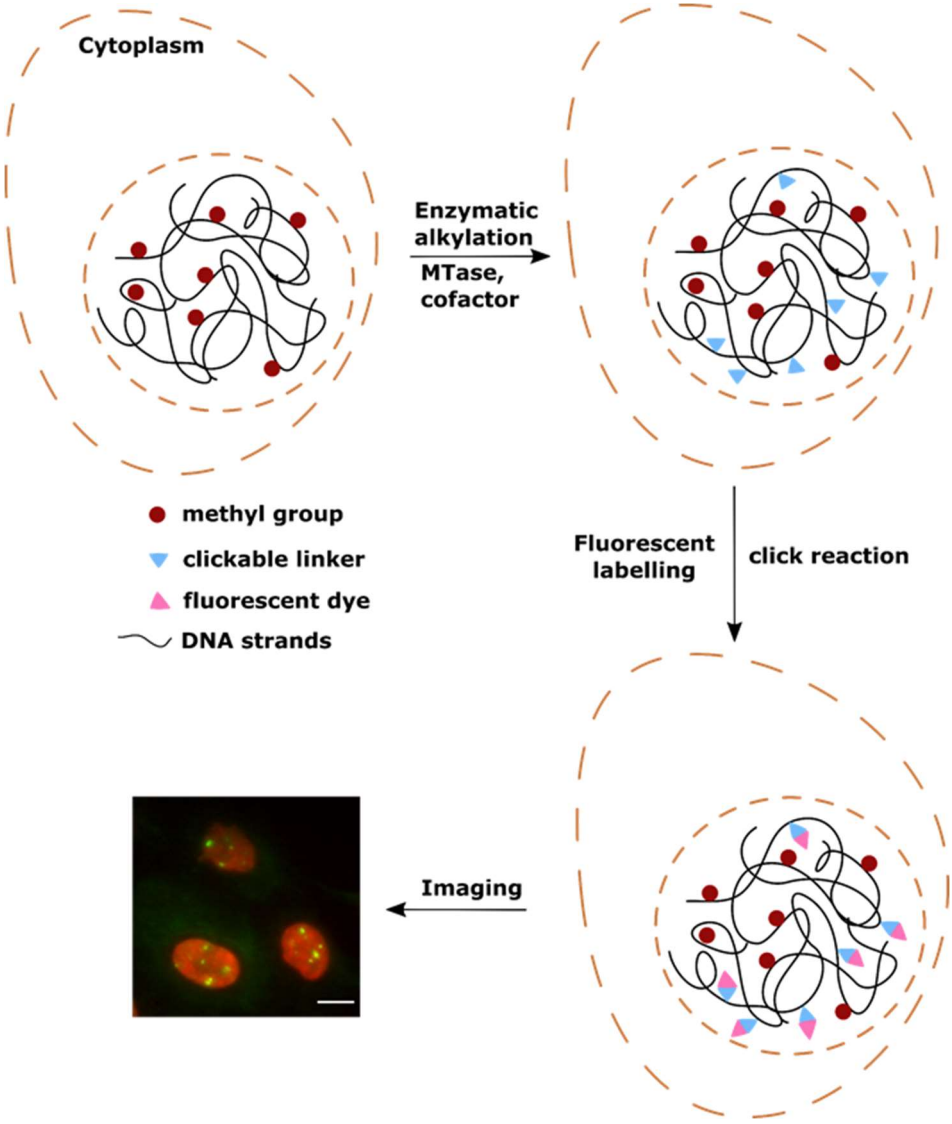


Fig. 49 A scheme of fluorescent labelling of unmethylated DNA in fixed cells. Widefield image of MCF10A cells labelled with M.TaqI enzyme and AdoHcy-6-yne. Red – DAPI staining, green – M.TaqI labelled DNA; scale bar: 10 μm.

II.2.3. Results

II.2.3.1. The choice of methyltransferases

In order to study the most studied epigenetic modification to the DNA in cancer, the enzyme targeting 5'-CG-3' sequence is needed. In humans, DNMT1 methyltransferase is responsible for maintenance methylation and modifies cytosine in the CG dinucleotide sequence. However, no synthetic cofactors bearing a functional transferable linker has been developed so far to be recognized by DNMT1. Fortunately, many bacterial MTases have been found to accommodate a range of modified AdoMet analogues and used to introduce various functionalities onto DNA.[150, 221, 251, 252] Here, we decided to employ two bacterial MTases: M.TaqI and a double mutant of M.MpeI (Q136A, N347A).[252] M.MpeI, similarly to DNMT1, recognises and labels cytosine residues within the 5'-CG-3' sequence. However, the enzyme needs to be engineered to catalyse DNA transalkylation from AdoMet analogues and a high concentration (around 100 μM) of a cofactor is usually required to fully modify DNA.[252] The second enzyme, M.TaqI MTase, modifies adenine in 5'-TCGA-3' sequence. Although the targeted base is adenine, the CG dinucleotide is contained within TCGA sequence and M.TaqI activity is blocked by the presence of m5C in the recognition site.[253] As a result of the low abundance of native adenine methylation in human DNA, which would also block the alkylation reaction, M.TaqI can be employed as a "CpG-methylation sensitive" enzyme and should allow getting of a subset of all unmethylated and accessible CG containing sites of the genome.

II.2.3.2. A distribution of M.TaqI target sites across the human genome

In order to be able to represent unmethylated CG sites by a subset of TCGA it is important to know how similar the distributions of both groups across the human

genome are. To map TCGA and CG content we used a MatLab code developed by Dr. Darren Smith which reads the provided DNA sequence and maps sites of choice. We have analysed each chromosome using the reference human genome available from NCBI database (version GRCh38.p12). Histograms of the number of TCGA or CG restriction sites per 50 kilobases were produced (Fig. 50).

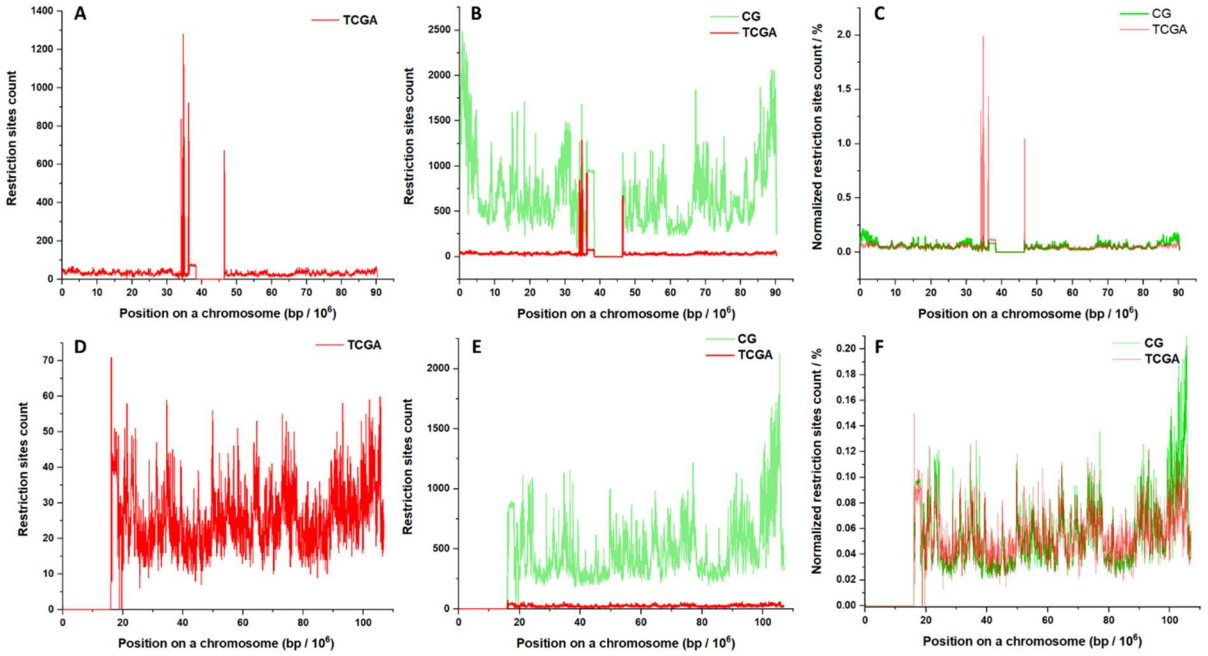


Fig. 50 A distribution of restriction sites on chromosome 16 (**A**,**B**,**C**) and chromosome 14 (**D**,**E**,**F**). (**A**) a distribution of TCGA sites on chromosome 16; (**B**) a comparison of distribution of CG (green) and TCGA (red) restriction sites on chromosome 16; (**C**) a comparison of a normalized distribution of CG (green) and TCGA (red) restriction sites on chromosome 16; (**D**) a distribution of TCGA sites on chromosome 14; (**E**) a comparison of distribution of CG (green) and TCGA (red) restriction sites on chromosome 14; (**F**) a comparison of a normalized distribution of CG (green) and TCGA (red) restriction sites on chromosome 14.

Two groups of the M.TaqI site distributions were identified (Fig. 50A,D). The first group has distinct localized peaks or group of peaks, with intensity up to 30 times greater than elsewhere on the same chromosome. Such peaks were observed on chromosomes 1, 2, 9, 10, 12, 16, 17, 20, 21, 22 and Y. In the regions without those peaks the distribution of M.TaqI target sites follows a similar pattern to the CG distribution (Fig. 50 E). However, the number of M.TaqI target sites within these small

regions can be quite significant i.e. up to 10% of all M.TaqI target motifs on chromosome 16 can be found in a region only 1.4% of its length, compared to 3.75% of CG content within the same region. The other group with mostly homogenous frequency of 30 - 40 sites per 50 kilobase pairs can be found on the rest of chromosomes: 3, 4, 5, 6, 7, 8, 11, 13, 14, 15, 18, 19 and X. In these regions, M.TaqI site density is around 10-fold lower than the frequency of CG motifs, which have a density of around 500 - 600 sites per 50 kilobase pairs with local variation of CpG islands (Fig. 50B,E). A normalized distribution of CG and M.TaqI target sites across the second group of chromosomes shows very similar patterns (Fig. 50 F).

Whether the distribution of CG motifs can be credibly represented by mapping of M.TaqI target sites depends heavily on the variation of the methylation level. Fig. 51 shows bisulfite sequencing of two cancerous cell lines, cervical – HeLa and lung – A549 (taken from ENCODE database), which maps m5C in CpG context, shows that methylation of a rich region of M.TaqI target sites on chromosome 16 is higher in cervical cancer cell line (HeLa) than in lung cancer cell line (A549). On the contrary, the methylation of a random M.TaqI target site poor region is high in both cell lines (Fig. 51). This could change depending on the presence of genes in these regions being differently active in both cell lines or other active elements. Therefore, this subset of the CG sequence could possibly be used to characterise and do comparative analysis of different cell lines e.g. identification of different cancer types or difference between healthy and cancer cells.

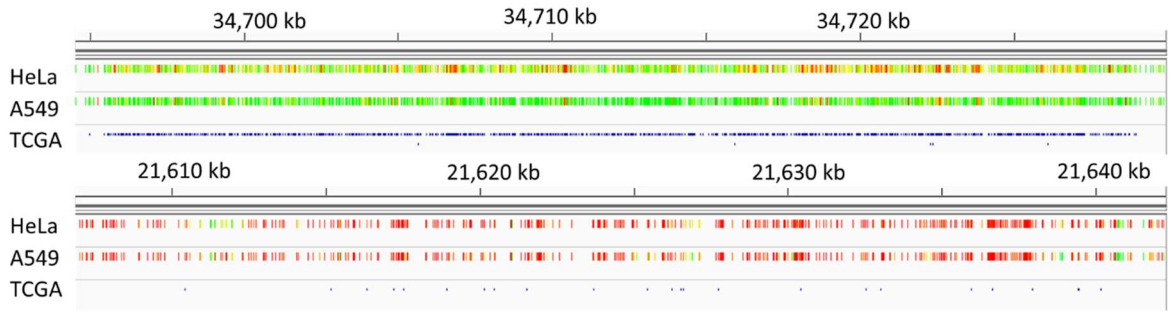


Fig. 51 A comparison of the CpG methylation of M.TaqI target sites rich region between cervical cancer cell line (HeLa) and lung cancer cell line (A549) in M.TaqI target sites rich (top) and poor (bottom) regions of chromosome 16. Red is high, orange medium and green low methylation. Blue shows M.TaqI target sites.

As the methylation status of regions containing a high M.TaqI site density can vary between different cell lines we compared these regions against existing sequencing data to identify any regulatory elements associated with them. The analysis showed that there is a variation of features identified in these regions such as long noncoding RNA, the vast majority is colocalized with DNase clusters (regions of accessible chromatin), transcription factor binding sites clusters and repetitive sequences such as SINE (Short Interspersed Nuclear Elements), LINE (Long Interspersed Nuclear Elements) and microsatellites (Fig. 52, see also Appendix B; for SINE and LINE see I.1.2.2).

Long noncoding RNAs (lncRNAs) are RNA transcripts of a length of over 200 nucleotides that do not code for functional proteins despite some of them being transcribed.[254] Although previously considered non-functional, recent studies showed a variety of processes that involve lncRNA molecules. It has been found that lncRNAs play a role in the regulation of chromatin structure by interactions with chromatin remodelling complexes, modulating gene expression[255] or the stabilisation of mRNA by blocking their degradation upon binding to mRNA.[256] lncRNAs are also considered to act as cancer biomarkers due to their tissue-specific expression and the role in affecting chromatin structure which can lead to development of cancer.[257, 258]

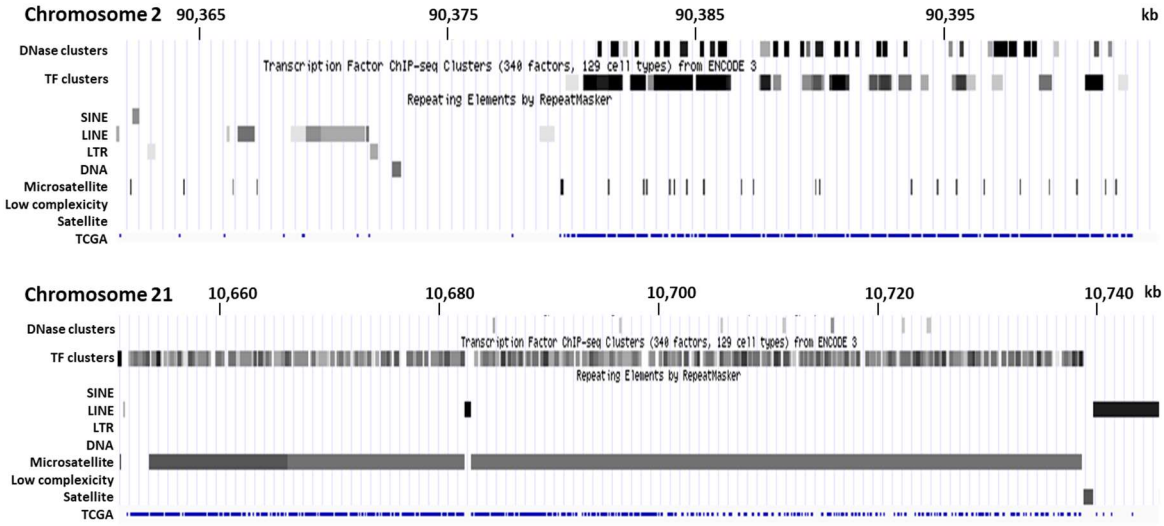


Fig. 52 Density maps of repetitive elements in rich regions of M.TaqI target sites on chromosomes 22 (top) and 21 (bottom). Density of M.TaqI target motifs in blue.

Another enriched genomic features identified in a high M.TaqI site density regions are microsatellites or Single Sequence Repeats (SSRs). They are 1 to 9 nucleotides long sequences occurring in tandem repeats across the whole genome. Due to the high level of mutation within them, microsatellites are used in genetic profiling and as a 'genetic fingerprint' in cancer diagnosis.[259] Such enrichment in clusters of transcription factor binding sites strongly suggests that these regions may play a role in regulation of gene expression. It has been found that about 33% of transcription factors are tissue specific[260] with many identified as oncogenes.[261]

The enrichment of DNase clusters in M.TaqI sites rich regions suggest their accessibility. Therefore, depending on the relative methylation state of regions containing a high M.TaqI site density to other parts of the genome, the labelling of genomic DNA with M.TaqI may preferentially provide the information on the activity (methylation level), organisation and function of regulatory elements such as lncRNAs, SINEs, LINEs or microsatellites. As expression of these elements can be tissue-

specific, the information on their methylation status could be used in the identification of diseases including cancer.

II.2.3.3. Optimisation of a method of genomic DNA labelling with M.TaqI in fixed cells

In the initial studies and method optimisation we decided to use HeLa cells due to their availability, high growth rate and straightforward culture. In order to access genomic DNA and conserve the structure of the chromatin, permeabilized and fixed cells were used in this experiment. The most commonly used fixation method using paraformaldehyde solution may introduce a number of modifications to nucleobases e.g. addition of formaldehyde to the nucleobases or form abasic sites.[262] As a result, the methyltransferase enzyme may not recognize the sequence or its activity may be hindered by addition of formaldehyde. Therefore, for fixation we chose a cold mixture of methanol/acetic acid (95:5 v/v) which causes dehydration and precipitation of biomolecules in cells and is more effective in the preservation of nucleic acids.[263] However, studies suggest that some variants of methanol-based fixatives may disrupt the chromatin structure.[264] Hence, obtained results need to be verified in the future and confronted with alternative fixation methods.

Firstly, we decided to establish an optimal cofactor and methyltransferase concentration. For labelling of cells with M.TaqI we chose the alkyne linker bearing cofactor - AdoHcy-6-yne - due to the efficiency of DNA transalkylation by M.TaqI using this cofactor (cofactor concentrations down to 3.12 µM *in vitro*) yielding full DNA protection (Fig. 53). M.TaqI activity is optimal at around 60°C. However, the incubation of cells at this temperature for an hour could have an impact on chromatin organisation. To minimize the possible changes to chromatin in cells we decided to carry out the

transalkylation reaction at 37°C. Thus, the M.TaqI activity was tested *in vitro* with AdoHcy-6-yne at 37°C (Fig. 53). Fortunately, pUC19 plasmid DNA is fully protected down to 3.125 µM AdoHcy-6-yne concentration.

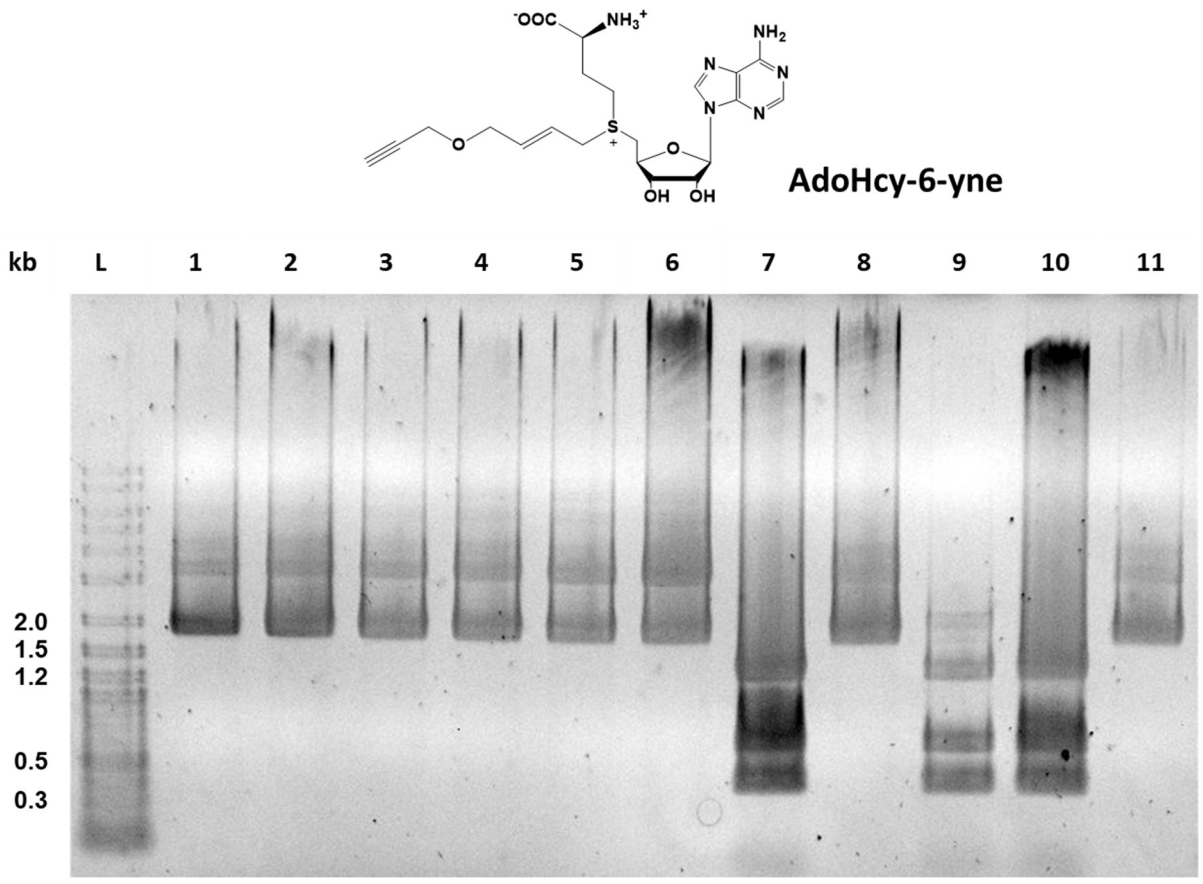


Fig. 53 Restriction assay with M.TaqI and AdoHcy-6-yne at 37°C. Lines: **1** – 100 µM AdoHcy-6-yne; **2** – 50 µM; **3** – 25 µM; **4** – 12.5 µM; **5** – 6.25 µM; **6** – 3.12 µM; **7** – No M.TaqI, 100 µM AdoHcy-6-yne; **8** – SAM control, **9** – No AdoHcy-6-yne; **10** – No M.TaqI, No AdoHcy-6-yne; **11** – No r.TaqI; M.TaqI concentration 14.5 µg/mL

Therefore we tested AdoHcy-6-yne in a range of concentrations (50, 12.5 and 3.12 µM) and with double the amount of M.TaqI compared to *in vitro* DNA labelling to ensure a maximum enzyme activity. Cells were labelled according to the method (see II.2.5) and images taken as a Z-stack.

The image of M.TaqI (-) control, in which cells were incubated with AdoHcy-6-yne at 3.12 μM, shows a high degree fluorescence emission across the cell (Fig. 54 left panel), which suggests the cofactor undergoes some off-target reaction within the cell. This non-specific fluorescent signal likely comes from a side reaction of the cofactor with nucleophilic groups present in the cell such as -NH2 or -S- (Fig. 54 right panel).

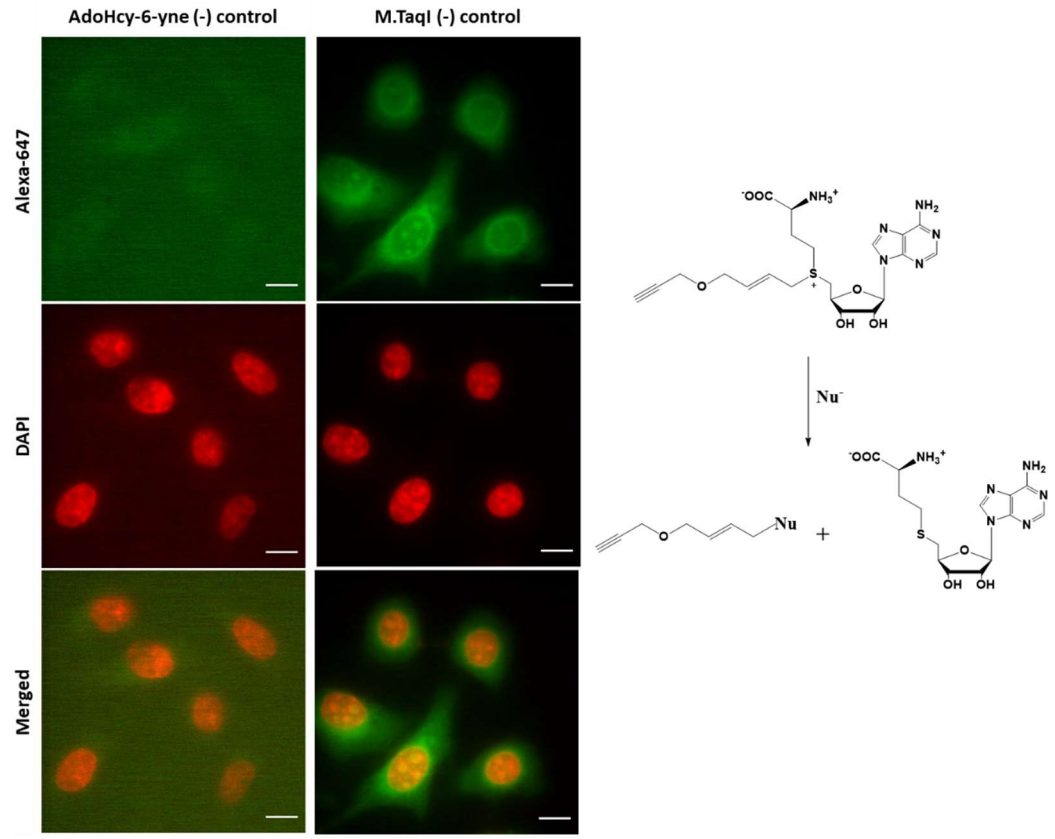


Fig. 54 Images of control samples of M.TaqI labelling with AdoHcy-6-yne cofactor (left panel). AdoHcy-6-yne (-) – cells incubated with M.TaqI 28.9 μg/mL in absence of AdoHcy-6-yne. M.TaqI (-) – cells were incubated with AdoHcy-6-yne (3.12 μM) in absence of M.TaqI. Scale bar: 10 μm. A single slice of a Z-stack shown. Side reaction of AdoHcy-6-yne cofactor with a nucleophile in cell (right).

As a result, a transferable linker is covalently bound to a host of biomolecules and fluorescently labelled during a click reaction.

The images at the highest tested concentration of the cofactor (Fig. 55 - 50 µM) show a similar non-specific fluorescence as M.TaqI (-) control. With decreasing concentration bright spots emerge within nuclei which could be in accordance with a previously identified regions containing a high M.TaqI site density (see II.2.3.2). They can be noticed at 12.5 µM cofactor concentration, yet the signal to “background” ratio

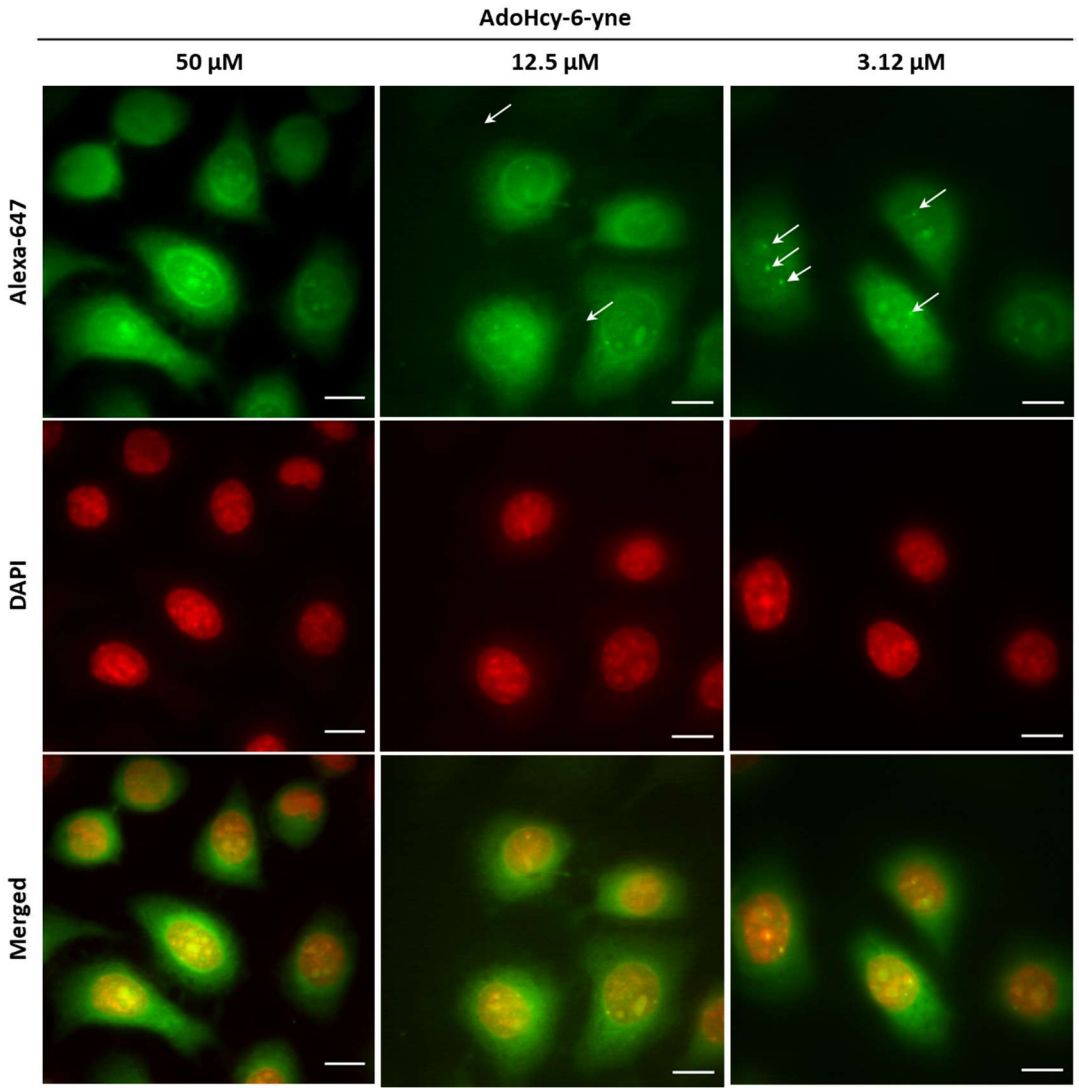


Fig. 55 Optimization of AdoHcy-6-yne concentration for unmethylated DNA labelling in cells. Alkylated DNA fluorescently labelled with Alexa-647-N_3, M.TaqI concentration 28.9 µg/mL. Scale bar: 10 µm. A single slice from a Z-stack shown.

is high. The lowest concentration of 3.12 µM shows the best results with spots clearly distinguishable (Fig. 53).

In the next step we decided to lower the concentration of the enzyme used. For this, a range of 28.9 – 3.62 µg/mL of M.TaqI was tested with a 3.12 µM concentration of the cofactor (Fig. 56). Images with the lowest concentration of M.TaqI (3.62 µg/mL) show that contrast between the clusters of genome-associated labels and the “background” is significantly worse than with higher concentrations of enzyme. Although the sample with 7.23 µg/mL of the enzyme does not seem to be much different from the second highest (14.45 µg/mL), it is in the lower end of efficiency (above 3.62 µg/mL). Thus, to

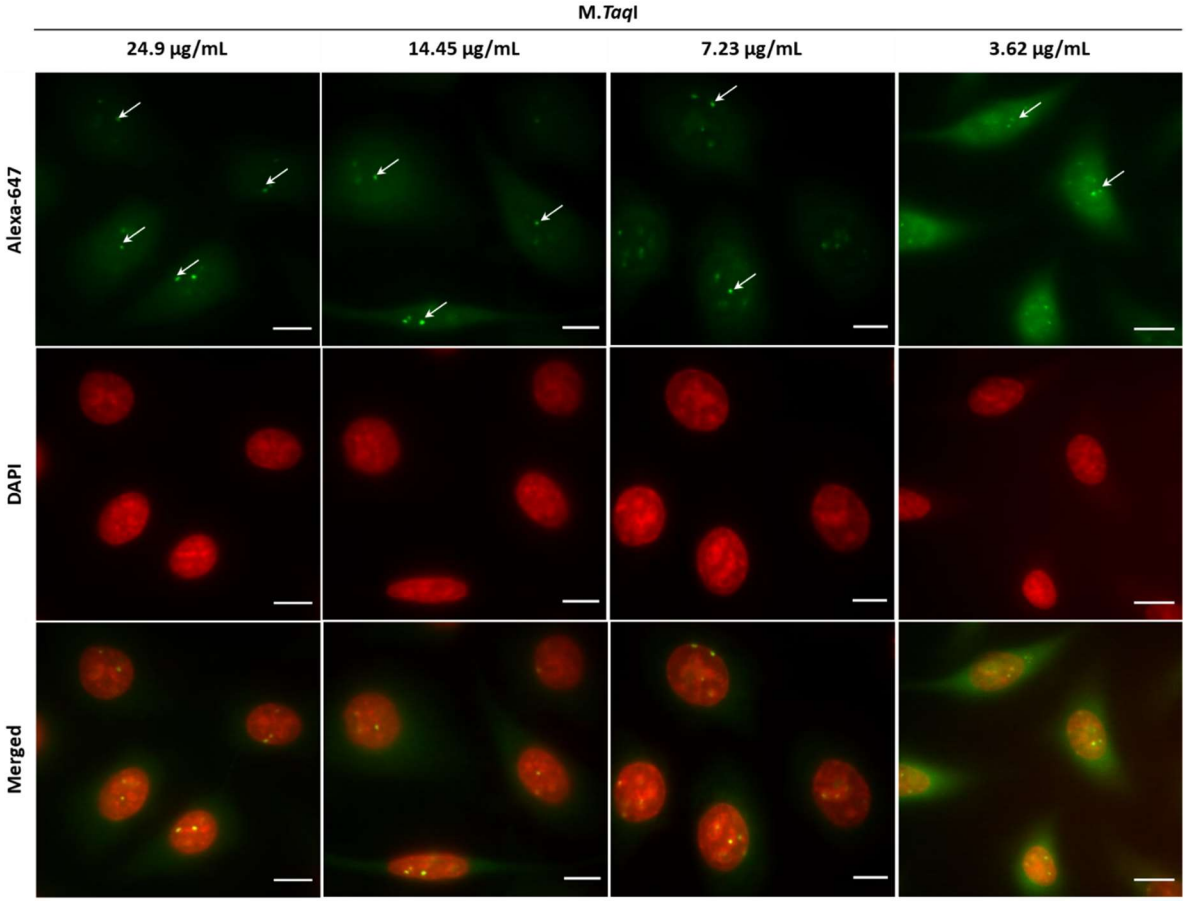


Fig. 56 Labelling of unmethylated DNA in fixed cell with serial dilutions of M.TaqI. Scale bar: 10 µm. A single slice from a Z-stack shown.

ensure full labelling and reproducibility of results a M.TaqI concentration of 14.45 µg/mL will be used in following experiments.

The results show that unmethylated M.TaqI target motifs are heterogeneously distributed across the nucleus. They tend to form clusters which can be found both in the outer and inner parts of a nucleus. What these clusters represent is so far unknown. Due to the uneven distribution of M.TaqI target motifs across the genome, clusters could represent highly dense M.TaqI target regions (Fig. 50). An alternative explanation is that observed clusters are induced in three dimensions by genomic condensation, such that the local concentration of M.TaqI target sites is relatively high.

In order to answer the question whether the observed clusters are previously identified highly dense M.TaqI target regions, further analysis with other methods, e.g. FISH (see II.2.1.1), is needed. Fluorescent oligonucleotides probes could be designed to target identified regions and subsequent imaging would reveal whether they colocalise with M.TaqI labelled regions of chromatin.

Once a method was optimised for labelling of fixed cells, we decided to use it and investigate an organisation of unmethylated M.TaqI target sites in healthy and cancerous cells from the same tissue.

II.2.3.4. The organisation of unmethylated M.TaqI target motifs in normal and malignant cells

For this study we chose to use two breast epithelial cell lines: MCF10A, as a model of non-malignant cell line, and MCF7, a breast cancer cell line. To investigate differences between the distribution and organisation of unmethylated M.TaqI target sites the cells

were fixed and labelled according to the general method (see II.2.5). Images were recorded as a Z-stack and foci counted manually (Fig. 57).

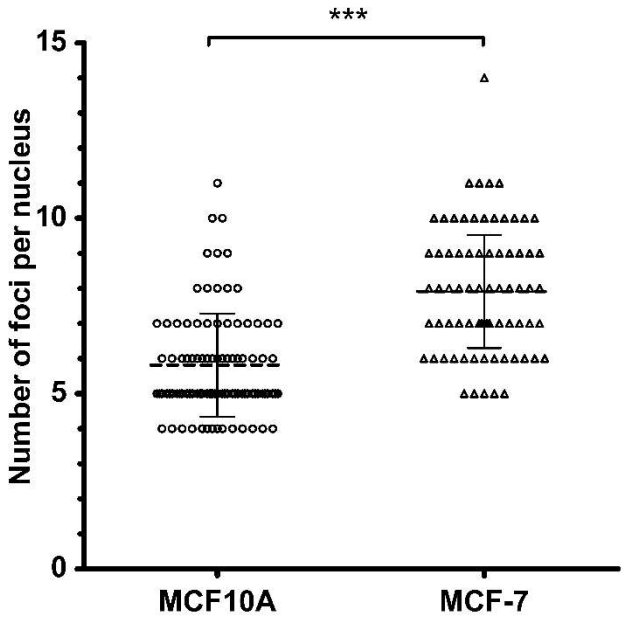


Fig. 57 A count of M.TaqI target foci number per nucleus. Data shown from three independent experiments. N = 100 nuclei per cell line counted. Means shown with dash lines, SD as error bars. Mann-Whitney test: $p < 0.0001$.

The number of foci per nucleus differs significantly between both cell lines. Almost a 50% more of foci can be found in cancerous MCF7 cells compared to their healthy counterparts (7.9 ± 1.6 and 5.8 ± 1.5 respectively). Reasons for this could be a difference in the methylation state of the genome, different organisation/accessibility of the genome in the two cell lines, or a varied number of copies of these DNA regions between the cell lines. The karyotype of cancer cells is usually more complex and differs from normal cells. It is characterized by an increased rate of chromosome gain/loss and mis-segregation. Therefore, compared to normal diploid somatic cells, consisting of 23 pairs of chromosomes, the karyotype of cancer cells can range from near-diploid (2N ± few) to even near-tetraploid (4N ± few).[265]

Compared to MCF10A,[266] the karyotype of MCF7[267] shows multiple additions of chromosomes (Fig. 58), which also includes chromosomes identified earlier with dense regions of M.TaqI target sites e.g. 1, 2, 12, 16, 17 or 20. Hence, if the signal is

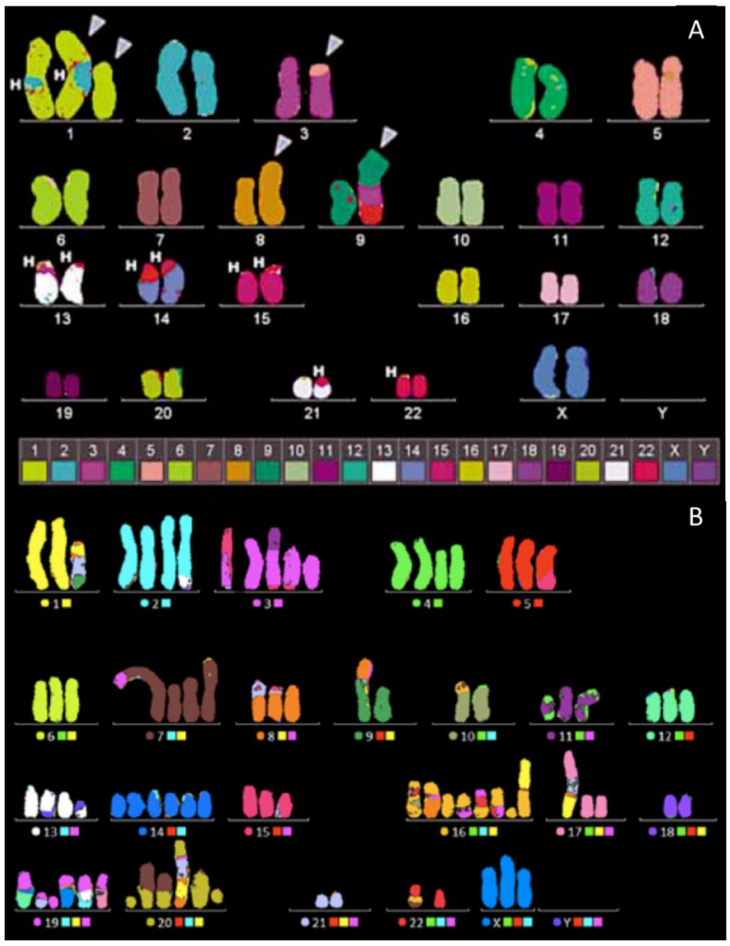


Fig. 58 M-FISH karyotype of a representative metaphase of MCF10A (**A**) and MCF7 (**B**) cells. Adapted from Rondón-Lagos M. et al, 2014 and Weise A. et al., 2003

dominated by those regions this would explain the increased number of foci per nucleus in MCF7 cells.

The counting of foci number is an oversimplified method for examining the differences in the (epi)genomes of the two cell lines and more subtle differences between the two cases can be observed. For example, MCF10A nuclei contain a lower number of foci per nucleus on average than MCF7 nuclei, but these foci are larger and more irregular

in shape. Such foci could be accumulations of several smaller labelling hotspots, that cannot be readily resolved (Fig. 59 A-C 1).

By contrast foci in MCF7 cells are more round, dot-like and usually can be counted separately even if close to each other (Fig. 59 D-F 2).

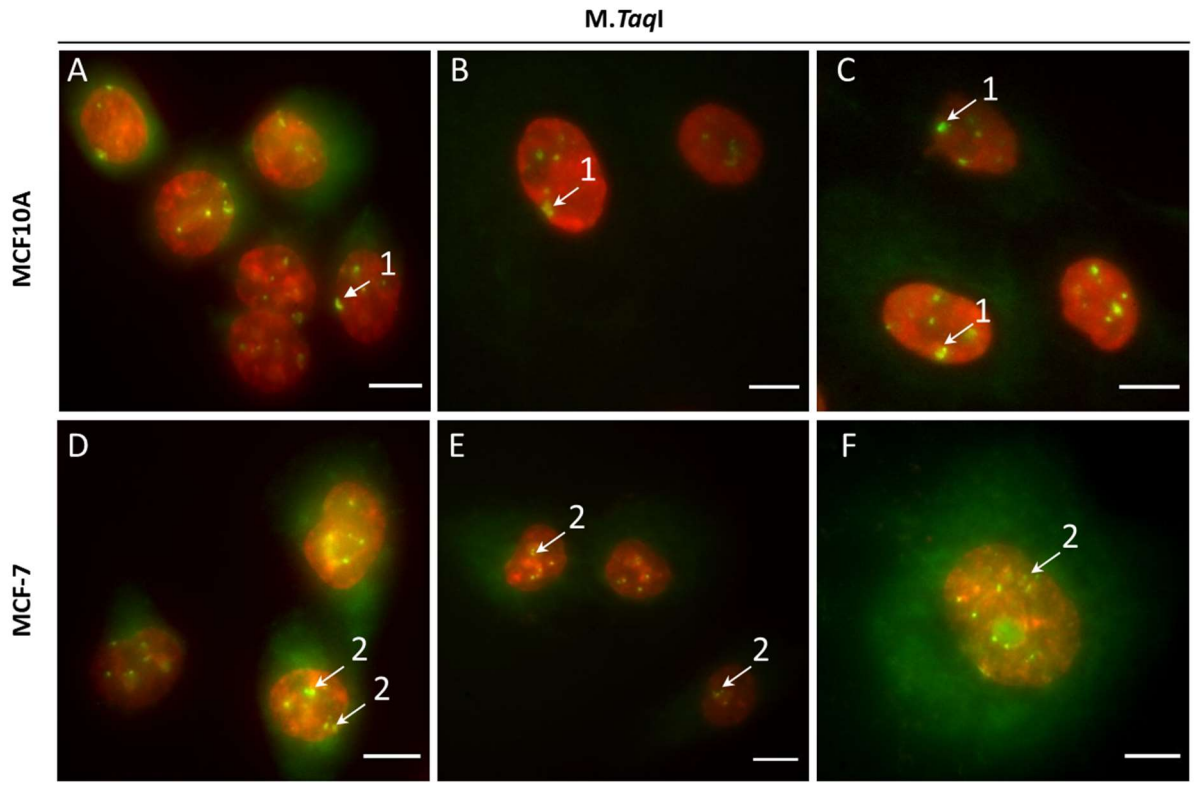


Fig. 59 A comparison of foci of M.TaqI target sites morphology in MCF10A and MCF7 cell lines. Images shown as a Z-projection of maximum intensity. Blue channel – DAPI; Alkylated DNA labelled with Alexa-647-N_3 (**A**, **D**) or Alexa-488-N_3 (**B**,**C**,**E**,**F**). Scale bar – 10 μm.

Foci of M.TaqI target sites can be found in different locations of a nucleus. An example of MCF7 nuclear foci are localized at the periphery of a nucleus (Fig. 60 A, D, G), in the inner part (Fig. 60 B, E, I) as well as the periphery of a nucleolus (Fig. 60 C, F, J). Linear profiles of intensities of DAPI and M.TaqI channels were plotted across each focus type (Fig. 60 K-M). DAPI is a nucleic acid stain preferentially binding to dense heterochromatic and A-T rich regions of the chromatin.[268] Dense chromatin is inaccessible, compact and it is typically composed of repetitive elements. Positive

correlation of increased intensity of DAPI and M.TaqI channels suggest that foci are preferentially located in such densely packed heterochromatic regions. This would be

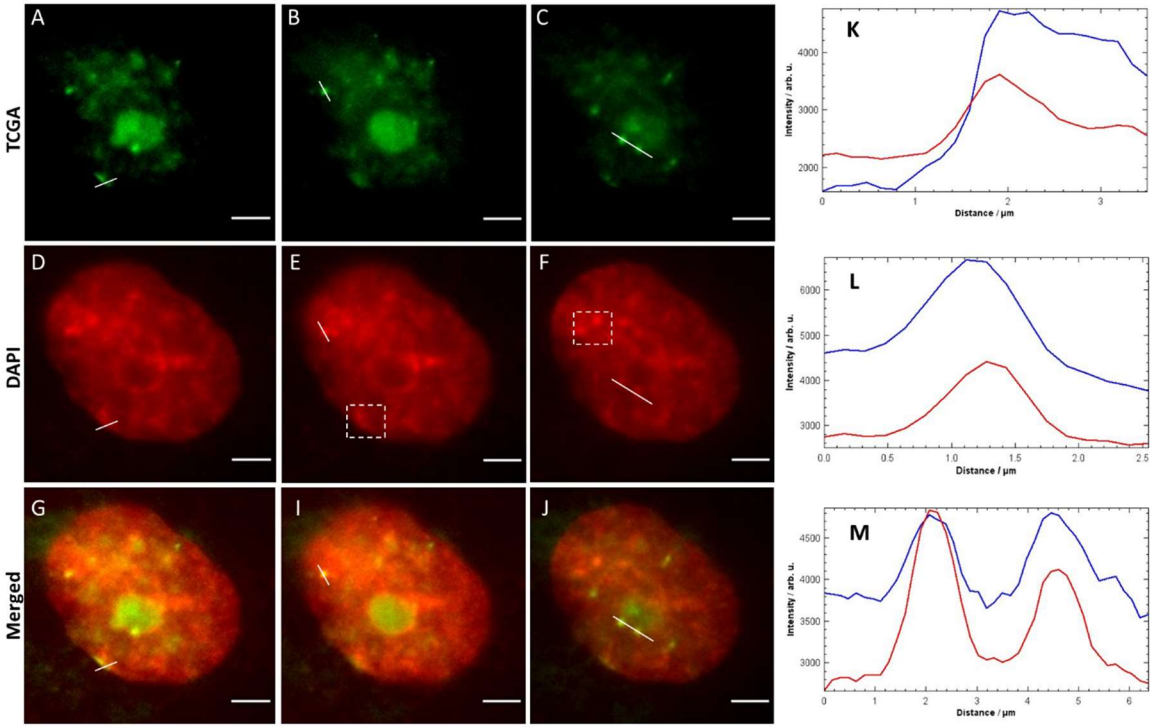


Fig. 60 The distribution of unmethylated foci of M.TaqI target sites across MCF7 cell nucleus (different Z-stack slices). **A**,**D**,**G** – peripheral location; **B**,**E**,**I** – inside the nucleus; **C**,**F**,**J** – two foci close to the nucleolus; **K**,**L**,**M** – intensity profiles of DAPI (blue) and M.TaqI target sites(red) in (K) peripheral location (G white line), (L) single spot (I white line), (M) two spots near the nucleolus (J white line). Scale bar: 5 µm

in agreement with high density M.TaqI target regions located in tandem repeat-rich regions. However, more insight is needed to support this conclusion.

II.2.3.5. Identification of fluorescent foci

To investigate the origin of observed foci in nuclei we decided to study their colocalization with other epigenetic modifications on histone proteins. One way to achieve this is to look for other imaging targets that should co-localise with the repetitive elements of the genome, in which we see large clusters on M.TaqI sites. To do this, the available ChIP-sequencing data for MCF7 cell line (for MCF10A no ChIP-

seq data was available) from the ENCODE database was used. ChIP-sequencing is a method of immunoprecipitation of chromatin fragments using an antibody against a histone mark of choice. Then DNA fragments, associated with the particular histone marker, from the captured chromatin are sequenced and aligned to the full genome sequence. Both gene activation marks such as H3K4me1 and repressive marks such as H3K9me3 were studied (Table 3). Out of the sequencing data available, for nine histone epigenetic marks only trimethylation of lysine 9 of histone 3 (H3K9me3) showed a significant enrichment in genomic regions we have identified as carrying high densities of M.TaqI sites (23 out of 27 identified regions).

	H3K4me1	H3K4me2	H3K4me3	H3K9me2	H3K9me3	H3K9ac	H3K27me3	H3K27ac	H3K20me1
High	-	-	-	-	10	-	-	-	-
Medium	-	-	-	-	7	-	-	-	-
Low	-	-	3	3	6	1	2	-	2
None	27	27	24	24	4	26	25	27	25

Table 3 An enrichment of epigenetic modifications of histone in dense regions of M.TaqI target sites in MCF7 cells. Blue columns – activation marks; red colulmns – repressive marks.

H3K9me3 is a heterochromatin mark and its signal in immunofluorescent studies is colocalized with DAPI staining.[269] However, in order to directly confirm co-occurrence of M.TaqI labelled clusters, the staining with H3K9me3 is needed in the future studies. Therefore, we decided to look into a relative localization of foci of M.TaqI target sites, two histone marks (activation – H3K27ac and repressive H3K27me3) and DAPI as a representation of heterochromatin in MCF10A and MCF7 cells. The cells were labelled according to the general method (see II.2.5). Linear profiles of the intensity across foci from three channels were taken and assigned into three groups depending on the overlap of the M.TaqI labelled DNA maximum with either: maximum (Fig. 61 I, J –

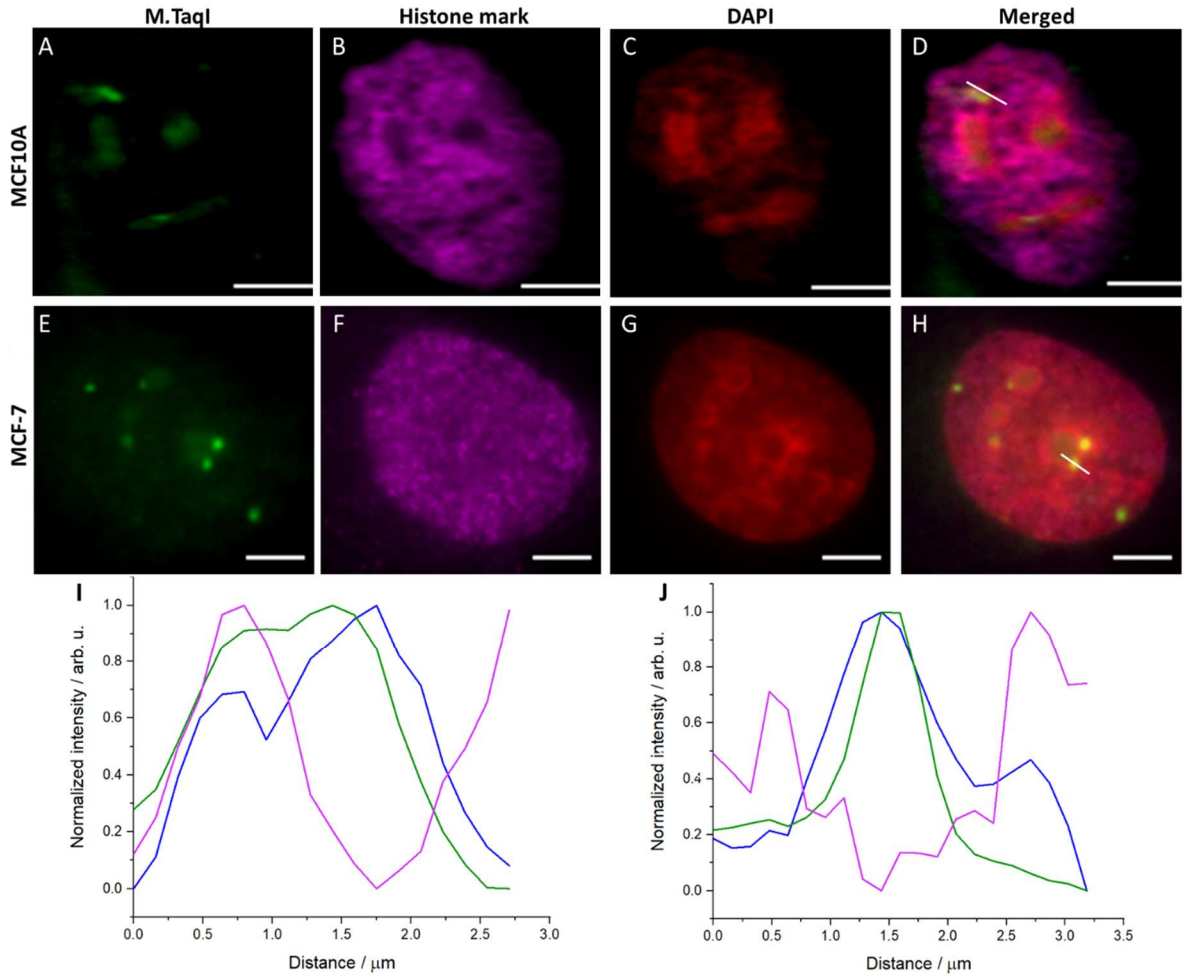


Fig. 61 A colocalization of M.TaqI labelled DNA foci, histone marks and DAPI signal. (**A**, **E**) – M.TaqI labelled DNA channel (Alexa488-N_3) of (A) MCF10A and (E) MCF7 cells; (**B**) – H3K27ac; (**C**, **G**) – DAPI; (**D**, **H**) – merged channels with a line indicating analysed regions (**I**, **J**) – normalized intensity profiles across the lines (D, H) respectively with green – foci of M.TaqI target sites, magenta – histone mark, blue – DAPI. Scale bar – 5 μm

M.TaqI and DAPI), intermidate (Fig. 61 I – M.TaqI and histone), minimum (Fig. 61 J – M.TaqI and histone) of other two fluorescence signals – DAPI and histone marker. A hundred foci were measured for each cell line and each histone mark (H3K27me3 and H3K27ac). For both cell lines and both chosen histone marks a vast majority of foci (86 - 94 %) are located where minima of histone marks signals occur (Fig. 62). In the images, spots of M.TaqI target sites seem to be located in pockets surrounded by increased signal from histone marks (Fig. 61). On the contrary, almost all (97 %) foci overlap with increased DAPI signal (Fig. 62). This is consistent with the theoretical

‘very low’ or ‘no enrichment’ of H3K27me3 and H3K27ac in dense regions of M.TaqI target sites. In addition, almost exclusive co-localisation of spots of M.TaqI target sites

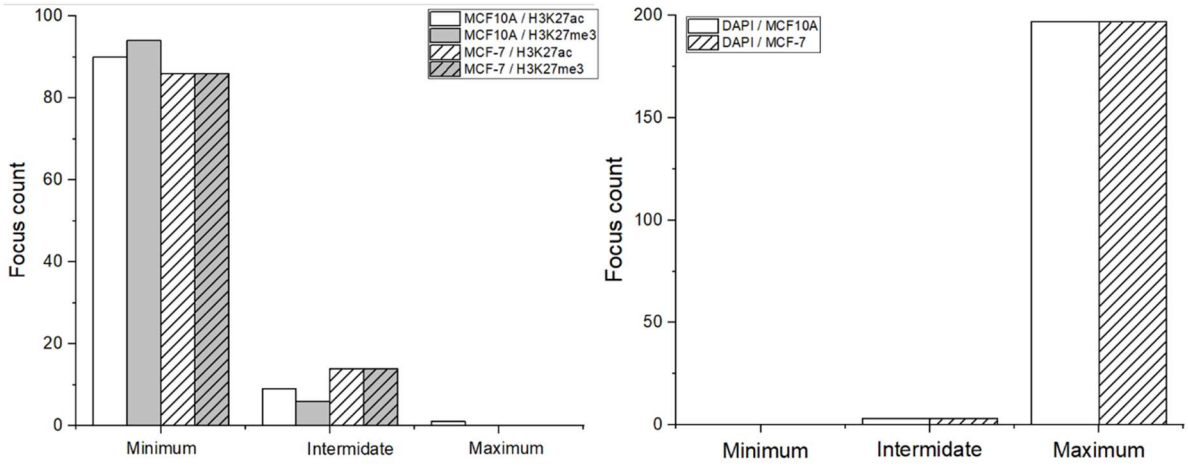


Fig. 62 A distribution of colocalization of foci of M.TaqI target sites signal with a maximum, minimum or partial overlap (intermidate) of histone marks (left) and DAPI (right) signal intensity.

with high intensity DAPI signal (equal to H3K9me3) would further suggest that observed foci are the previously identified dense regions of M.TaqI target sites in the genome.

We have developed and preliminarily validated a method for labelling unmethylated regions of DNA in fixed cell using the M.TaqI MTase (with TCGA sequence recognition). We now aim to examine the organisation of unmethylated CG dinucleotide sequences.

II.2.3.6. Labelling of unmethylated CG regions in fixed cells

The methylation of cytosine within the CpG dinucleotide is a very important epigenetic modification and plays a role in many processes such as gene expression and chromatin organisation. Its malfunction can lead to many diseases including cancer. Therefore, we decided to test our method for labelling of unmethylated CpG sites. We used an engineered bacterial MTase enzyme (M.*Mpe*I Q136A, N347A) that, like the

native human DNMT1 enzyme, methylates cytosine within CpG sequence. Unlike DNMT1, M.*Mpe*I displays good activity with synthetic cofactors bearing functional groups as a transferable linker. First, we tested the previously used AdoMet analogue AdoHcy-6-yne, which bears a linker with terminal alkyne group suitable for subsequent labelling using azide-alkyne cycloaddition. A restriction assay was carried out in a range of AdoHcy-6-yne concentration of 500 – 7.812 μM (Fig. 63 B). Many methyltransferase enzyme have a respective restriction enzyme recognizing the same sequence. Unfortunately, there is no restriction enzyme recognizing CG sequence and therefore *Hpa*II enzyme recognizing 5'-CCGG-3' was used. The circular plasmid DNA (pUC19) containing 12 *Hpa*II recognition sites was used in the assay. Therefore, the restriction pattern is more complicated than with 4 sites in case of M.TaqI. A MatLab code developed by Dr. Darren Smith was used to predict the gel bands pattern at various restrictions levels, 0 – 100 % restriction efficiency depending on the methylation by M.MpeI (Fig. 63 A). From the gel, pUC19 is fully methylated by M.MpeI when AdoMet is used for 500 μM and 125 μM concentrations (Fig. 63 B lanes 5 and 6). However, for 31.25 μM AdoMet restriction efficiency is 10% and increases to 90-100% for 7.812 μM AdoMet (Fig. 63 B lanes 7 and 8 respectively). This shows that M.MpeI is at least 10 times less active with its native cofactor than M.TaqI is with synthetically modified AdoHcy-6-yne at 3.12 μM (Fig. 53). Unfortunately, no activity of M.MpeI was observed with AdoHcy-6-yne even at the highest 500 μM concentration.

As AdoHcy-6-yne is not a suitable cofactor analogue for by M.MpeI, we decided to use a synthetic cofactor bearing an azide as a terminal group in the transferable linker – AdoHcy-6-N_3 (Fig. 64). This cofactor was synthesized and characterized by Dr. Elodie Jagu. During the purification on semi-preparative HPLC, both isomers of this cofactor are collected separately. The activity of both isomers was tested in a restriction assay in a concentration range 250 – 62.5 µM of AdoHcy-6-N_3 (Fig. 64). The alkylation efficiency for all the concentrations of both isomers was below 50 % even at 250 µM. As we tested previously with AdoHcy-6-yne a high concentration of the cofactor

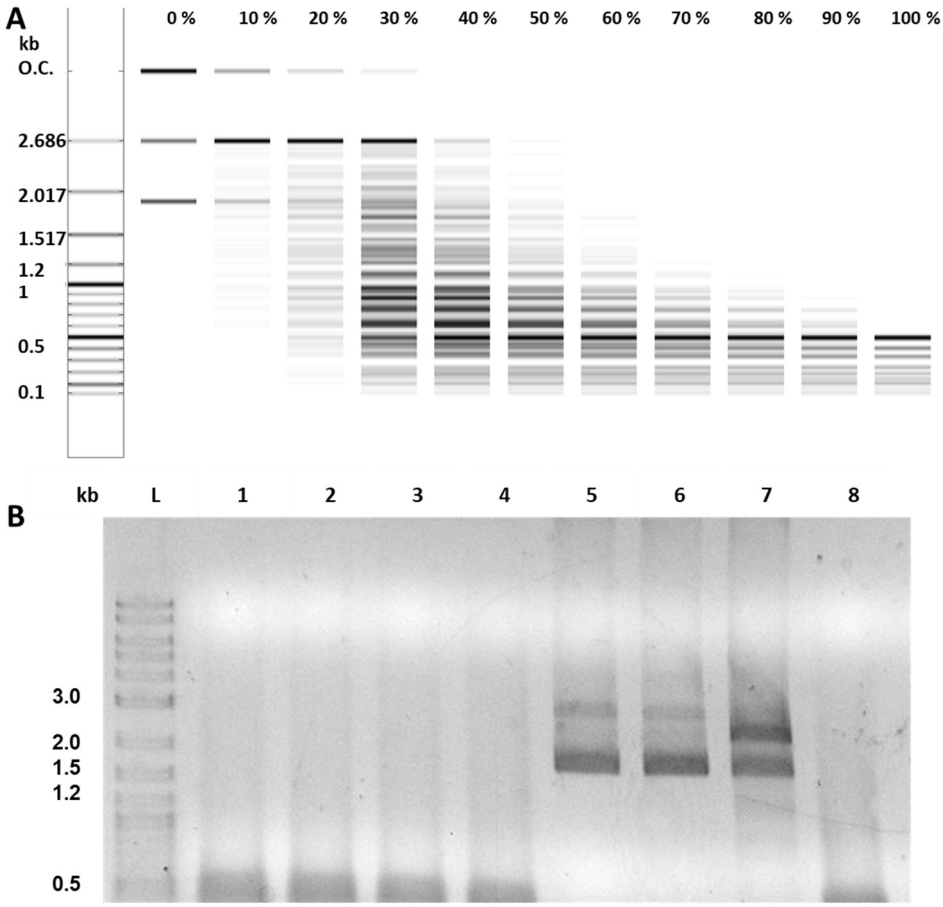


Fig. 63 A simulation of the restriction pattern for pUC19 DNA and HpaII restriction enzyme (0 – 100 % restriction efficiency) (**A**), an agarose gel of restriction assay for AdoHcy-6-yne and M.MpeI (**B**). (**B**) **1** – AdoHcy-6-yne 500µM; **2** – AdoHcy-6-yne 125 µM; **3** – 31.2 µM; **4** – 7.81 µM; **5** – AdoMet 500µM; **6** – AdoMet 125 µM; **7** – AdoMet 31.25 µM; **8** – AdoMet 7.81 µM. M.MpeI – 460 µg/mL

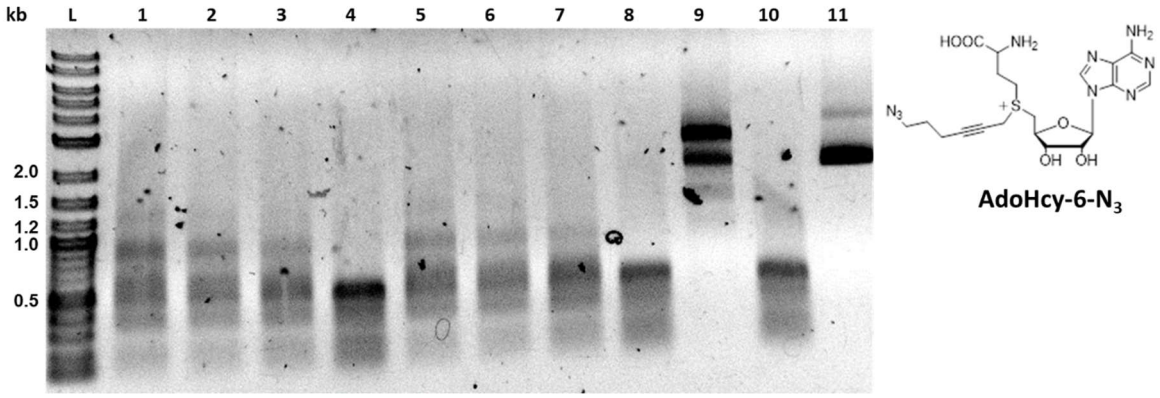


Fig. 64 An agarose gel of restriction assay for AdoHcy-6-N_3 and M.MpeI. 1 – AdoHcy-6-N_3_I 250 μM; 2 – 125 μM; 3 – 62.5 μM; 4 – AdoHcy-6-N_3_I 250 μM, No M.MpeI; 5 – AdoHcy-6-N_3_II 250 μM; 6 – 125 μM; 7 – 62.5 μM; 8 – AdoHcy-6-N_3_I 250 μM, No M.MpeI; 9 – AdoMet 62.5 μM; 10 – No cofactor; 11 – AdoMet 62.5 μM, No HpaII.. M.MpeI – 138 μg/mL

creates a “background” fluorescence signal which overshadows the signal from labelled DNA (Fig. 55). Hence, the lowest tested concentration (62.5 μM) was used in the cell labelling experiment. Unfortunately, a background signal from non-specific transfer of the azide-terminated linker onto nucleophiles in the cells was higher and homogenous (Fig. 65). No distinguishable features in nuclei can be observed which suggests that 62.5 μM concentration of the cofactor is too high for visualization of unmethylated CpG sites using this approach.

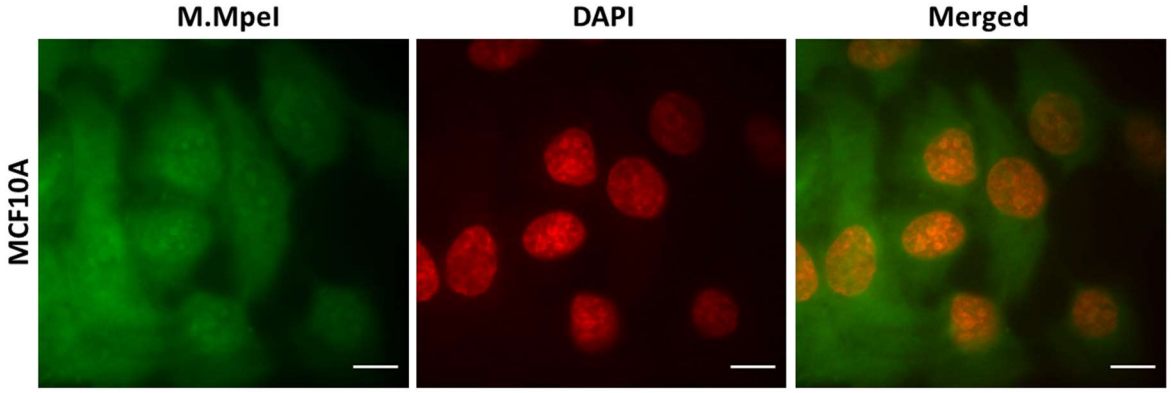


Fig. 65 Widefield images of genomic DNA labelling with M.MpeI in fixed MCF10A cells. AdoHcy-6-N_3 62.5 μM; scale bar – 10 μm

In order to extract the genomic information out of the non-specific alkylation signal and to study complex structures of foci of M.TaqI target sites another method is needed.

II.2.4. Conclusions

We have developed a method to site-selectively fluorescently label and visualise the architecture of accessible and unmethylated DNA in fixed cells. This was done using M.TaqI methyltransferase enzyme (targeting 5'-TCGA-3' sequence) and a synthetic analogue of AdoMet cofactor (AdoHcy-6-yne) bearing a clickable alkyne linker. The method was optimised in order to minimise required concentration of both enzyme and cofactor, and to improve the signal to noise ratio of labelled DNA. Widefield microscopy was applied to reveal organisation of unmethylated M.TaqI target sites in cells into foci.

The analysis of the theoretical distribution of M.TaqI recognition sequences (5'-TCGA-3') across the genome revealed that this sequence is not uniformly and randomly distributed but rather occurs more frequently in specific loci. These clusters likely contain repetitive, sequences including microsatellites which are used in cancer diagnosis as well being rich in transcription factors binding sites.

The comparative analysis of number of foci per nucleus was done on two breast epithelial cell lines (healthy model MCF10A and cancer MCF7 cells). This revealed that on average two more foci can be found in cancerous cells. However, due to the difference in number of copies chromosomes between tested cell lines it was not clear whether the increased number was caused by the difference in methylation of these regions or simply by the greater number of copies of particular chromosomes in cancer cells. It was also found that observed foci colocalise with regions of high DNA density (based on DAPI staining). In addition, most of the foci are found in the regions poor in two tested histone marks marking active (H3K27ac) and repressed chromatin (H3K27me3) which is in agreement with theoretical analysis. However in order to

confirm the origin of observed signal, more studies are needed including FISH experiments or IF against H3K9me3 .

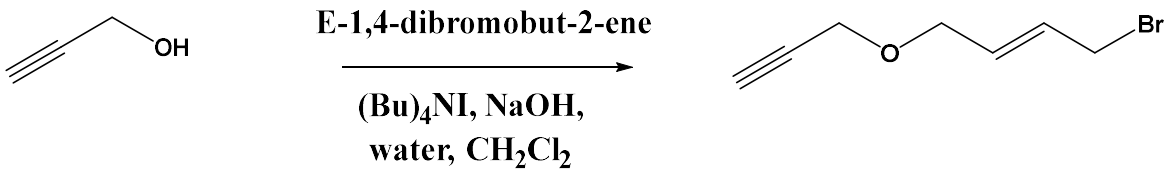
This method was also tested with bacterial M.MpeI enzyme which has the same recognition pattern as human DNMT1 enzyme (5'-CG-3') and AdoHcy-6-N_3 cofactor. Using widefield microscopy we were not able to resolve any signal from site-selectively labelled DNA. This was likely due to higher concentration of cofactor required, compared to MTaqI studies, which resulted in higher "background" fluorescence signal from a non-specific transfer of a clickable linker from the cofactor onto nucleophiles in cells.

II.2.5. Materials and Methods

Widefield microscopy

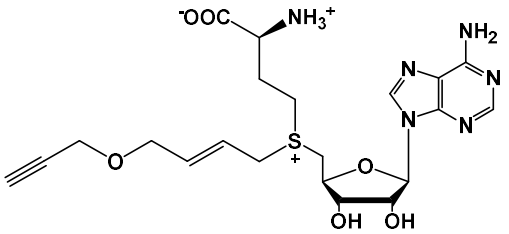
Imaging of fixed cells was performed on an ASI RAMM microscope, equipped with a Nikon 100x 1.45 NA TIRF objective. Illumination was from a 100mW OBIS CW lasers via a quad-band dichroic mirror (405/488/561/640) and images were collected using an Evolve Delta EM-CCD camera via a quad-band emission filter (Semrock, 432/515/595/730nm). Micromanager was used to control the system and scan the sample.

Synthesis of (E)-1-bromo-4-(prop-2-ynyloxy)but-2-ene



(E)-1-bromo-4-(prop-2-ynyloxy)but-2-ene was synthesised according to the method published before with small adjustments.[217] Briefly, a solution of propargyl alcohol (300 mg, 5.35 mmol), tetrabutylammonium iodide (197.7 mg, 0.1 eq.) and crushed NaOH (2.14 g, 10 eq.) in 12 mL of dichloromethane and 1.5 mL of water was stirred for 5 minutes. Then, a solution of (E)-1,4-dibromobut-2-ene (2.29 g, 2 eq.) in 3 mL of dichloromethane was added at once on ice and stirred overnight at room temperature. The mixture was diluted with 15 mL of ethyl acetate and aqueous layer separated. Organic layer was washed twice with 15 mL of water, saturated solution of $Na_2S_2O_3$ (15 mL), brine and dried over Na_2SO_4. The solvent was removed under reduced pressure and the crude was purified by flash chromatography (DCM/hexane 1.5:8.5 v/v until (E)-1,4,-dibromobut-2-ene starts eluting then hexane/EtOAc 5:1). The solvents were removed under reduced pressure yielding pale yellow oil (756 mg, 4.02 mmol, 75%). NMR spectrum in accordance to literature results was obtained. TOF MS ES (+) $[M+H]^+$ calcd. 186.9759, found 186.9751.

Synthesis of AdoHcy-6-yne



To a solution of AdoHcy (25 mg, 0.065 mmol) in $HCOOH/CH_3COOH$ (1:1, 500 µL), (E)-1,4,-dibromobut-2-ene (500 mg, 40.9 eq.) was added slowly at 0°C. Then, a solution of $AgClO_4$ (16 mg, 1.2 eq.) in 100 µL of $HCOOH/CH_3COOH$ (1:1) mixture was added. The reaction mixture was stirred for 16 hours at 30°C in dark. 2 mL of water was added,

AgBr precipitate was filtered off using 0.45 µm membrane filter and (E)-1,4,-dibromobut-2-ene was extracted with diethyl ether. Then, traces of organic solvent were removed under reduced pressure. The solution was purified by preparative RP-HPLC (3-100% MeOH in 20 mM ammonium formate buffer pH 3.5 over 60 min) and the collected fraction was lyophilized. The product (0.0369 mmol, 56.7%) was stored at -20°C in 0.1% formic acid solution. TOF MS ES (+) m/z M^+ calcd: 493.1869, found: 493.1877.

Restriction assay using M.TaqI, pUC19 and AdoHcy-6-yne

For pUC19 restriction assay, the master mix (120 µL) was created on ice of 50 ng/µL pUC19, 14.5 µg/mL M.TaqI in NEB Cutsmart buffer pH = 8.5. The master mix was split into 9 vials (20 µL in first and 10 µL to 2-6, 8, 9 and 11). To the first vial, AdoHcy-6-yne in 0.1% formic acid solution was added to reach the concentration of 100 µM. Then a half of the sample 1 was transferred into sample 2. It was continued until sample 6. AdoMet solution was added to samples 8 and 11 to reach 100 µM concentration. Two additional controls (10 µL each) were prepared with the same concentrations except the corresponding components were omitted: control of AdoHcy-6-yne at 100 µM with M.TaqI omitted (number 7) and with both M.TaqI and AdoHcy-6-yne omitted (number 10). All the samples were incubated for 1 hour at 50°C (or 37°C). 0.5 µL r.TaqI was added into all samples except one of the AdoMet controls (number 11) and then samples were incubated for 1 hour at 50°C. Next, 0.5 µL of Proteinase K (20 mg/mL) was added to each sample and all were incubated for 1 hour at 50°C. Samples were run on 1% agarose gel (120V for 40 minutes).

Restriction assay using M.MpeI (Q136A, N347A), pUC19 and AdoHcy-6-N_3

For pUC19 restriction assay with M.MpeI and AdoHcy-6-N_3, the master mix (255 µL) was created on ice of 25 ng/µL pUC19, 138 µg/mL M.MpeI (Q136A, N347A) in NEB Cutsmart buffer pH = 8.5. The master mix was split into 8 vials 30 µL each (1-3, 5-7, 9 and 11). AdoHcy-6-N_3 solution in 0.1% formic acid was added to samples 1-3 (isomer I) and 5-7 (isomer II) to reach desired concentration 250-62.5 µM. AdoMet solution was added to samples 9 and 11 to reach 62.5 µM concentration. Three additional controls (30 µL each) were prepared with the same concentrations except the corresponding components were omitted: control of AdoHcy-6-N_3 (isomer I or II) at 250 µM with M.MpeI omitted (4 and 8) and with both M.MpeI and AdoHcy-6-N_3 omitted (number 10). All the samples were incubated for 1 hour at 37°C. Then, 4 µL of proteinase K solution (20 mg/mL) was added to each sample and then samples were incubated for 1 hour at 50°C. The samples were purified using GenElute PCR Clean-Up kit (Sigma Aldrich) following manufacturer's protocol with elution into 50 µL of water heated up to 50°C. To 20 µL of each sample 2 µL of Tango buffer (10x) (Thermo Fisher) and 0.3 µL of HpaII (10 U/µL) (Thermo Fisher) were added and samples were incubated for 1 hour at 37°C. Next, 0.3 µL of proteinase K (20 mg/mL) was added to all samples and samples were incubated for 1 hour at 50°C. Samples were run on 1% agarose gel (120V for 40 minutes).

Cell Culture

MCF7 and HeLa cells were cultured in Dulbecco's Modified Eagle's Media (DMEM) (Sigma) supplemented with 10% fetal bovine serum (FBS) (Gibco) and 5mM glutamine (Sigma). MCF10A cells were cultured in DMEM/F-12 supplemented with 25% horse serum, 500 ng/mL hydrocortisone (Sigma), 10 µg/mL insulin (Sigma), 20 ng/mL EGF

(Miltenyi) and 100 ng/mL cholera toxin (Sigma). All cell lines were grown in 5 U/mL penicillin and 1mg/mL streptomycin.

DNA alkylation in cells using bacterial methyltransferase enzymes

Cells were grown overnight in 24-well plate on poly-L-lysine coated Ø1.3 cm glass cover slip (40000 cells/well seeded). Cells were then washed twice with PBS and fixed with solution of methanol and glacial acetic acid (95:5 v/v) for 10 minutes at -20°C.

Cells were then washed twice with PBS and stored at 4°C until used in the labelling experiment. Cells were incubated with a solution of methyltransferase enzyme (M.TaqI at 8.7 µg/mL, M.MpeI at 115 µg/mL final concentration), cofactor (AdoHcy-6-yne at 3 µM, MTC-5 at 65 µM final concentration) in 1x CutSmart® buffer at 37°C for 1 hour. Samples were then washed twice with PBS and incubated with 4 mM $CuSO_4$, 2 µM fluorescent dye and 100 mM sodium ascorbate solution in TBS at room temperature for 30 minutes in dark followed by washing twice with PBST and once with PBS.

For immunofluorescence, after labelling procedure as above, cells were blocked in 10% FBS in PBST for 30 min at room temperature. Staining with primary antibody Anti-Histone H3 (tri methyl K4) antibody (Abcam) (1:2400), Histone H3K27me3 antibody (pAb) (Active Motif) (1:1000) or Histone H3K27ac (Acetyl Lys27) antibody (GeneTex) (1:1000) for 1 hour at room temperature in blocking solution, washed three times with PBST and incubated with respective secondary antibody (Donkey anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 568, ThermoFisher or Donkey anti-Goat IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 633, ThermoFisher) (1:2000) in blocking solution for an hour at RT. Washed with PBST twice and once with PBS.

For widefield microscopy on the slides cells were labelled as above and stained with 2 μM DAPI in PBS for 5 minutes in dark and mounted onto glass slide with ProLong™ Gold Antifade mountant.

II.3. Application of expansion microscopy for studying the organisation of unmethylated genomic DNA

II.3.1. Expansion microscopy

To tackle a high non-specific background signal and low image resolution we applied expansion microscopy (ExM). The method, developed in 2015 by Edward Boyden's group, aims to physically expand the specimen and therefore improve image resolution using a conventional microscope. It uses an *in situ* synthesized swellable polymer network to expand the sample.[270] Two monomers, common acrylamide and superabsorbent sodium acrylamide are used alongside N-N'-methylenebisacrylamide as a crosslinker in free radical polymerization reaction. An accelerator, usually tetramethylethylenediamine (TEMED), and initiator, ammonium persulfate (APS), are used to trigger the reaction. The formed gel can be dialysed against a low salt concentration solution e.g. deionized water. As a result, the gel expands up to 4.5-fold in each dimension, a 91-fold increase of a volume (Fig. 66 A). Gels with an expansion factor of 10x have also been reported although this employs a relatively challenging sample preparation.[270, 271] The biological sample, such as cells or tissue slices, is incorporated into polymer mesh during the polymerization reaction via small molecule anchors, which are covalently attached to proteins, DNA or RNA.[272, 273] For fluorescence imaging a biological target i.e. DNA, RNA or protein, can be fluorescently labelled either pre- or post-gelation. Subsequently, the sample is digested with a non-specific protease leaving a matrix of fluorescent dyes anchored to the polymer network.

In the next step, the gel is expanded usually in water (Fig. 66 D). In this work, gels with ~4x expansion factor were used (Fig. 66 B,C). Although ExM was developed very recently, a number of modified ExM protocols have been published for various applications such as ExFISH for RNA imaging,[273, 274] proExM for imaging of specific proteins using fluorescent antibodies in both single cells and tissues.[272, 275, 276]

The expansion microscopy provides a tool to achieve a resolution below the diffraction limit of light (~300nm) using a conventional widefield microscope. A resolution of 60 – 70 nm can be achieved using 4.5x gels (~300/4.5 nm = ~67 nm).[270] In addition, after digestion with protease and expansion in water, the gel matrix contains mostly water and some DNA and RNA molecules and, hence, is optically transparent. Therefore, the samples can be imaged in 3D on a widefield microscope, which greatly increases the speed of volumetric imaging compared to high resolution methods such as STORM, PALM or STED which require long acquisition times in order to achieve sub-diffraction limit resolution.[208, 209, 277, 278]

ExM is not without limitations. The main issue is that some fluorescent dyes e.g. cyanine based Alexa647 are degraded during the polymerization reaction and therefore are not suitable for ExM. Even the most robust dyes such as Alexa488 and Atto647N typically retain up to only 70% of their initial fluorescence, post-polymerisation.[272]

This compatibility issue can sometimes be solved by staining after the polymerisation step such that none of the labels are damaged by the polymerisation reaction. Transparency of the gel, which is a great advantage, can also cause some difficulties. A lack of natural reference points makes imaging of the same part of the sample or the

same cell challenging. This can be facilitated by cutting a gel into a easily recognizable and characteristic shape.

As the gel consists mainly of water, it may collapse or gradually release water causing the sample to move during imaging. To tackle this, a poly-lysine coated (positively charged) surface can be used to limit movement of the negatively charged gel. The gel can also be embedded in agarose and immersed in water during imaging, which also stabilizes the gel.[279]

Here, we apply ExM to increase the imaging resolution and to remove the majority of proteins which were non-specifically labelled by cofactor in the cells (Fig. 66 D).

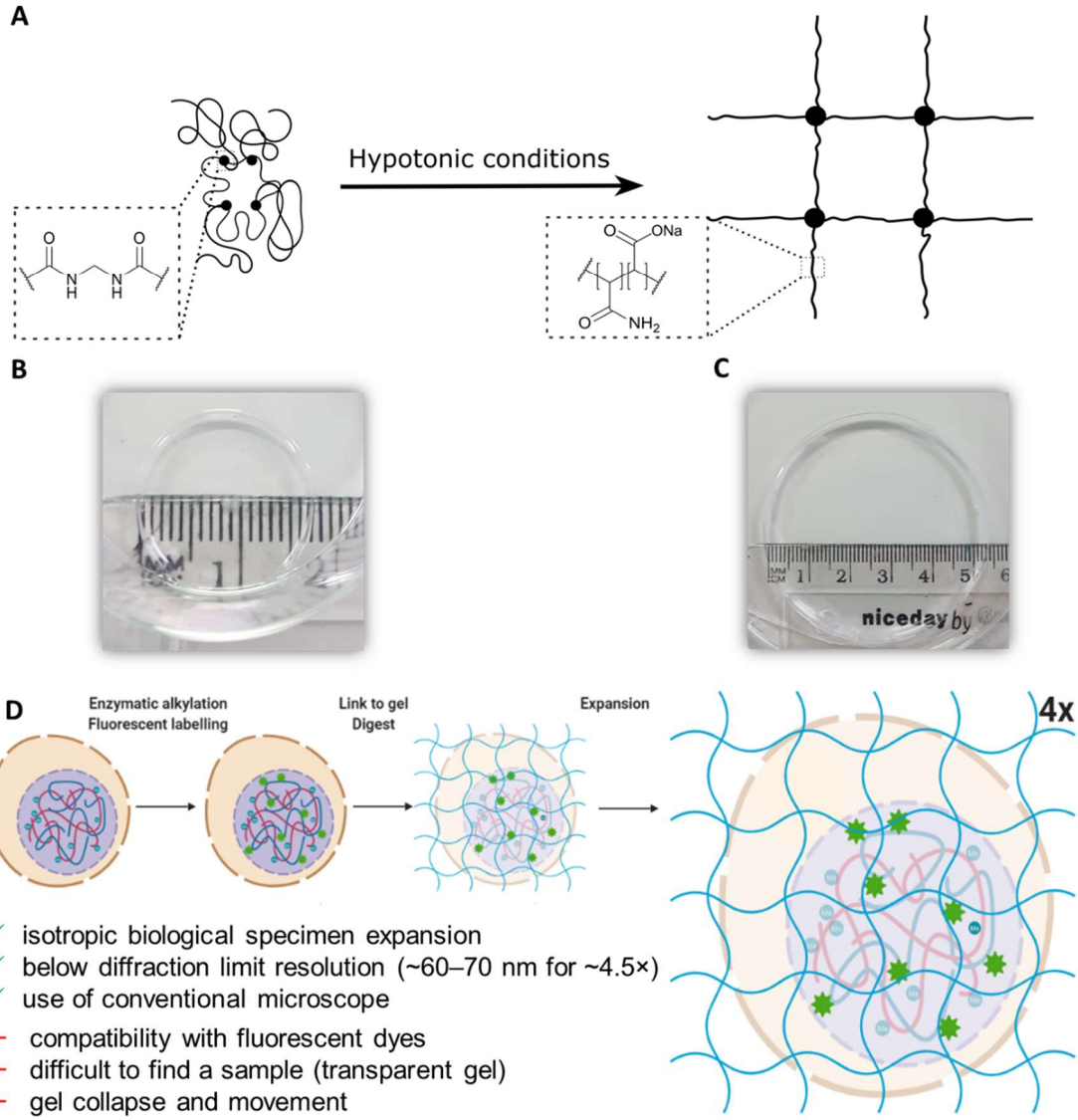


Fig. 66 A scheme of gel expansion mechanism (**A**), images of a gel post-digestion and post-expansion (**B**,**C**) and an overview of the expansion of biological sample . (A) The crosslinker shown as dots, polymer scaffold (from acrylamide and sodium acrylate) as lines. A tangled network after polymerization (left) and after expansion in hypotonic solution i.e. H_2O. Negatively charged acrylate groups repel causing gel expansion. (B) A gel after digestion (diameter 1.7 cm, expansion factor 1.3x) and the gel after dialysis in deionized water (diameter ~5.2 cm, expansion factor ~4x) (C). Expansion microscopy of the unmethylated genomic DNA. First, DNA is alkylated using MTase enzyme and a synthetic cofactor, and fluorescently labelled. Next, the DNA is linked to the gel via anchor and proteins digested. A transparent gel with a matrix of fluorescent dyes and DNA is expanded in deionized water to reach ~4x expansion factor (64-fold volumetric expansion).

II.3.2. Results

II.3.2.1. An overview of the method

The alkylation of unmethylated genomic DNA using M.TaqI MTase can be imaged in fixed cells under a widefield fluorescence microscope yielding bright clusters distributed across the nucleus. However, the complex structure of some of the foci

cannot be resolved using standard, diffraction-limited microscopy. Furthermore, lower intensity signal from regions that are less dense in M.TaqI labelling sites can be lost due to the signal resulting from non-specific reaction of the synthetic cofactor. This effect is particularly pronounced when a higher concentration of the cofactor is used, as is the case for the less active M.MpeI MTase. In this case only a pancellular signal is observed.

The new labelling procedure applying ExM consists of a few steps as follows (Fig. 66 D). Similarly to the previously applied protocol, cells are grown on a poly-lysine coated glass coverslip (Ø13 mm), and fixed with cold methanol/acetic acid (95:5 v/v). Subsequently, cells are labelled with a synthetic cofactor (AdoHcy-6-yne or AdoHcy-6-N_3) and MTase (M.TaqI or M.MpeI) of choice followed by fluorescent labelling using the azide-alkyne cycloaddition reaction. In the next step, cells are incubated with the DNA anchor – LabelX – bearing a vinyl group for the polymerization reaction. The gelation is carried out on a glass slide covered in parafilm layer which prevents the gel from sticking to the glass. 90 µL of mixed monomers solution, TEMED and APS is pipetted onto the slide and a glass cover slip is gently placed onto the droplet with cells facing down (Fig. 67). The gelation reaction is carried out at 37°C for 1-2 hours followed by digestion with Proteinase K overnight at RT or up to 6 hours at 37°C. After the digestion step the coverslip detaches from the gel or can be easily peeled off. The gel itself expands slightly during this digestion step (~1.3x up to Ø 1.7 cm) (Fig. 66 B). The gel is then placed in deionized water. Water is exchanged every 30-40 minutes until the gel stops expanding, usually 3-4 exchanges of the water are needed. Eventually, the gel expands ~4-fold in each dimension reaching ~5.2 cm in diameter (Fig. 66 C).

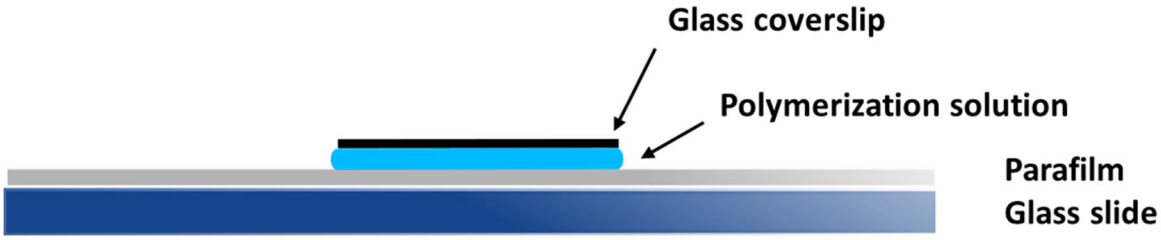


Fig. 67 Scheme of a gelation reaction setup

The resulting gels were placed on the glass bottom dish and imaged on the widefield microscope. In case of the expanded gel, to limit the sample movement, the gel can be embedded in the 1-2% agarose.

II.3.2.2. Expansion microscopy to study the organisation of unmethylated M.TaqI target motifs of genomic DNA in the nucleus

First, the gels with M.TaqI labelled MCF10A cells were used to develop the protocol. A control sample with no enzyme shows that after the digestion step, no background can be detected apart from within the nucleoli (Fig. 68 A). This contrast with the previous attempt, and suggests that the background signal from the cells on coverslips was indeed from proteins that were non-specifically labelled, both in the cytoplasm and nuclear region, and that these have been successfully removed by digestion with proteinase K, that is performed as part of the expansion process.

The remaining homogenous signal in nucleoli suggests that fluorescent dye or cofactor conjugated to fluorophore is trapped in nucleoli and surrounded by the intact DNA (Fig. 68 A arrow 1). However, in the green channel, the area occupied by the nucleus can be distinguished in M.TaqI (+) samples after expansion (Fig. 69 A-E). This homogenous nuclear signal is likely a result of M.TaqI-directed DNA labelling across the genome. In the image of the pre-expanded sample labelled with M.TaqI, clear foci can be detected alongside stained nucleoli (Fig. 68 B-D). Compared to the pre-

expanded sample, foci in the post-expansion sample show more complex structures, such as the sickle-like shaped feature in Fig. 69 A,B arrow 2. Furthermore, the selected

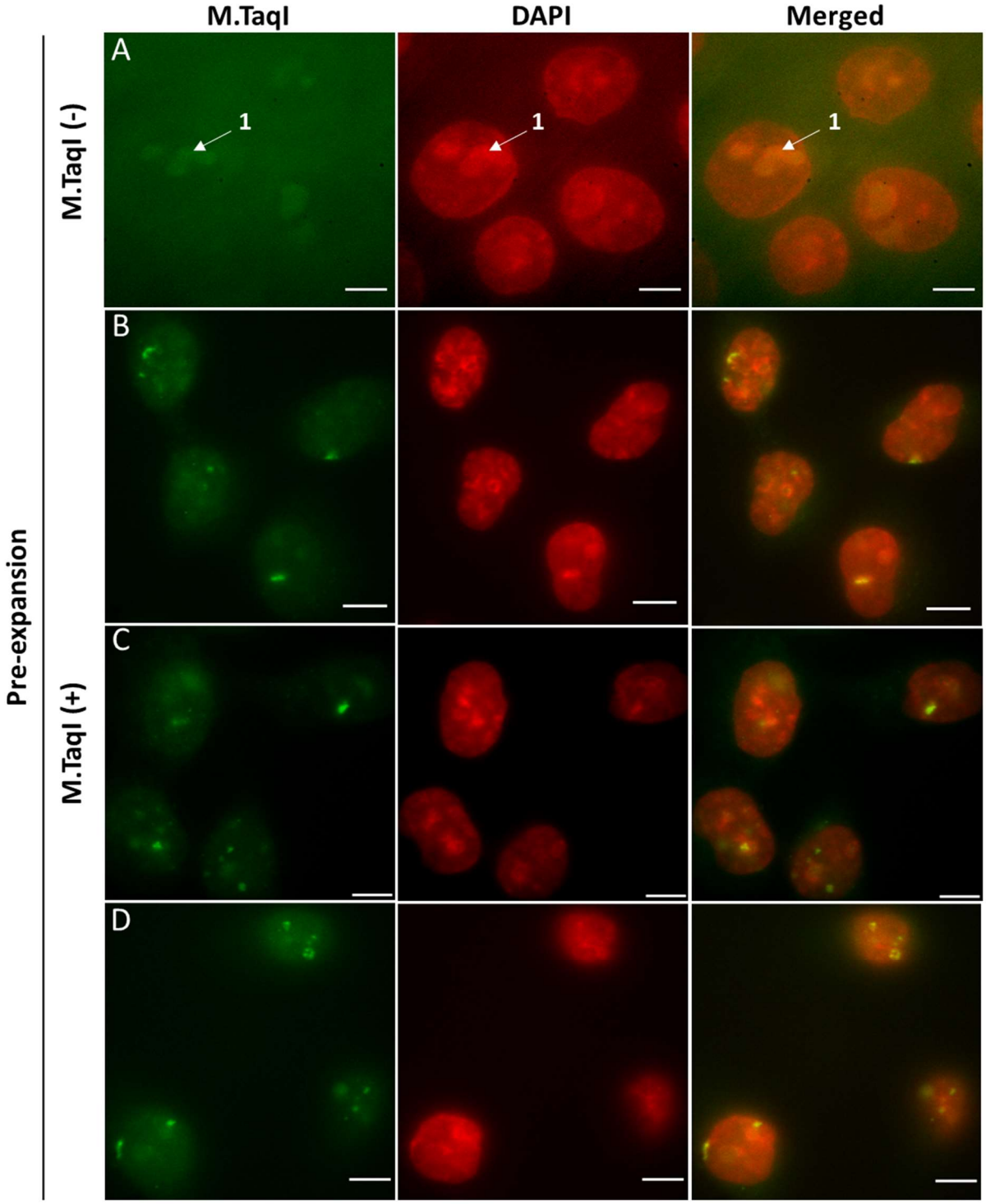


Fig. 68 Widefield images of expansion microscopy gels of MCF10A cells: pre-expansion (**A**) control with no M.TaqI, (**B-D**) M.TaqI and AdoHcy-6-yne labelled. Scale bar = 10 μm, single slices of Z-stack presented.

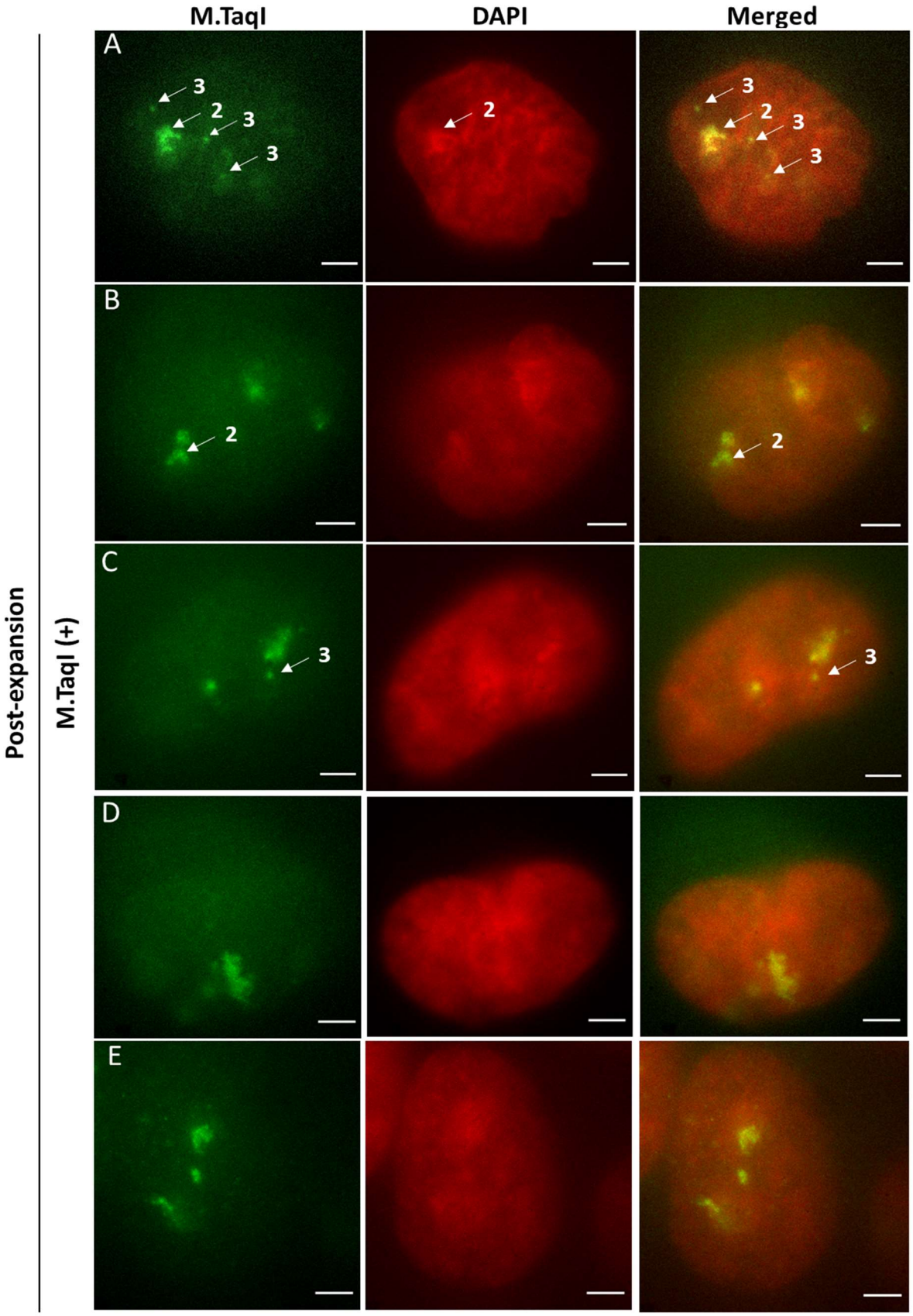


Fig. 69 Widefield images of expanded microscopy gels of MCF10A cells (**A-E**) M.TaqI and AdoHcy-6-yne labelled. Scale bar = 10 µm, single slices of Z-stack presented.

cluster overlays with similar structures in the DAPI channel which is in agreement with previous results (Fig. 62). After the background was significantly reduced by ExM, smaller and lower intensity clusters are visible in the expanded sample (Fig. 69 A,C arrow 3). Although, in the case of M.TaqI labelled cells, the issue with high background was mainly resolved the images are blurry due to the fluorescence from out-of-focus planes. This is more pronounced when the fluorescence signal in the Z-direction of the sickle-like cluster of M.TaqI target sites is presented (Fig. 70). Here, there is the

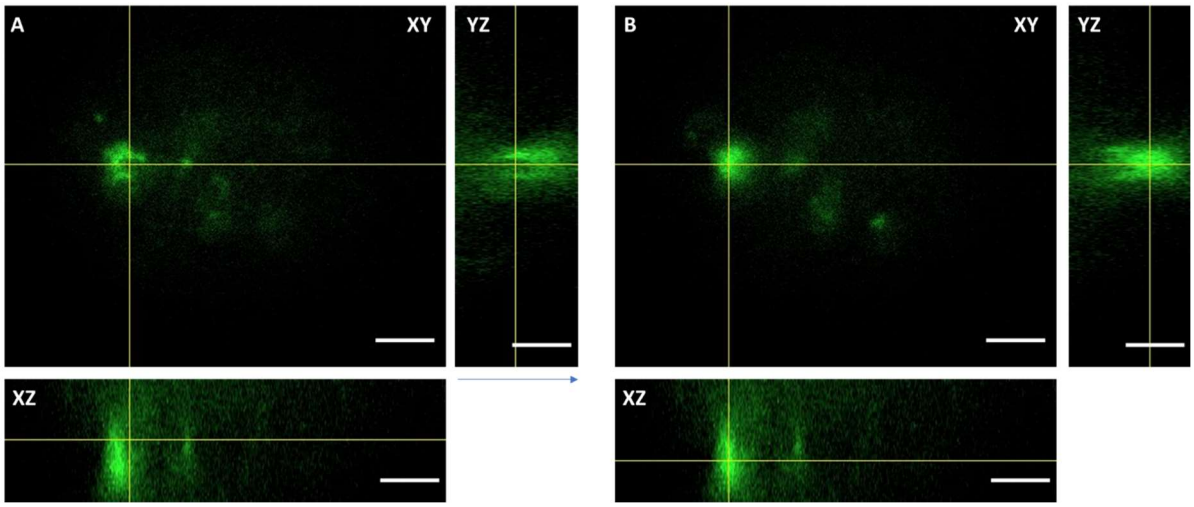


Fig. 70 Orthogonal views of clusters of M.TaqI target in expanded gel in two separate Z positions (**A**) bottom of the sample and (**B**) top of the sample. Scale bars = 10 µm, Z step = 0.5 µm. Contrast adjusted to expose the selected cluster.

suggestion that several sub-structures are present yet the out-of-focus blurring makes resolving them a challenge (Fig. 70 A). This structure, from YZ profile, appears to be around 17 µm in height, in the longest part, stretching across the majority of the thickness of the nucleus (Fig. 70 A, YZ). Therefore, from the image many possible structures of the cluster can be proposed. This includes a single object with wide oval top and sickle-like bottom or a cluster of structures that are separated in Z-axis but due to the background fluorescence from other planes appear to be a one large structure.

To test if the resolution both planar and axial can be improved, the deconvolution algorithm was applied. The structure of such clusters may differ between the cells

depending on the methylation level in the cell or the type of the cell. Hence, a better identification of the structure may provide important information and allow a characterisation of the cells.

II.3.2.3. Application of deconvolution algorithm to widefield images of ExM samples

The deconvolution algorithm uses the point spread function (PSF) of the optical setup to predict the contribution of the signal from out-of-focus planes and remove it from the in focus plane or volume, thereby improving the sharpness and resolution of the image (see I.3.3).

We decided to apply this method to improve the resolution of the widefield images of expanded M.TaqI labelled cells. For this purpose we decided to implement two plugins in *Fiji* software: *Diffraction PSF 3D* plugin to predict theoretically PSF for both channels used (405 nm and 488 nm) and *Iterative Deconvolve 3D* to perform deconvolution.[280] In order to automate the processing I wrote a script (see Appendix B). Deconvolution was performed with maximum 100 iterations and terminated when the difference of the mean of fluorescence in the image between iteration steps is less than 0.01%. Processing of a single channel of a Z-stack image, which usually contains only one nucleus, takes between 2-5 minutes depending on the number of iterations required.

The algorithm was applied to the Z-stack image of the same sample presented in Fig. 70. The same cluster was compared both before and after deconvolution (Fig. 71). Firstly, it can be seen that the blur was decreased by the deconvolution and structural features of the genome are more sharp in both the DAPI and M.TaqI channels (Fig. 71 A, B, E, F). The extent of the blur reduction is more noticeable in DAPI channel due to

the larger portion of the sample being labelled (nuclear DNA) compared to regions of M.TaqI target sites. The resolution in the XY plane is improved in both channels with a much more pronounced improvement in the DAPI channel resolution where two intensity peaks emerged in the line profile at a distance of 5 μm apart (Fig. 71 C, G). This provides more information on the structures visualized and more details on the morphology can be observed including more distinguishable foci and separation of the smaller, bottom part of the zoomed cluster (Fig. 71 F). Also, it seems that the overall shape of the nucleus is lost in the M.TaqI channel after deconvolution. This might be just a result of the out-of-focus fluorescence being removed/decreased hence the remaining signal is more difficult to distinguish from the background outside the nucleus (Fig. 71 E, F). However, it needs to be taken into account that some information might have been removed hence any analysis of these regions of low density of M.TaqI target sites might be misleading. In addition, a darker area around the bright cluster in M.TaqI channel is observed, which is a common deconvolution artifact and can be a source of a loss of information (Fig. 71 F1).

Axially, in both XZ and YZ planes the blur was also reduced and resolution improved which can be observed in the YZ intensity profiles (Fig. 71 D, H) of the feature highlighted with a white line (Fig. 71 A, B, E, F; YZ). Thus, now the height of the highlighted structure is ~15 μm after deconvolution in both channels compared to 18-20 μm depending on the channel which is a change of 16.7-25 % (Fig. 71 D, H). Now the ‘U’ axial profile of the cluster is more clear which is especially visible in the XZ plane of M.TaqI channel (Fig. 71 F; XZ).

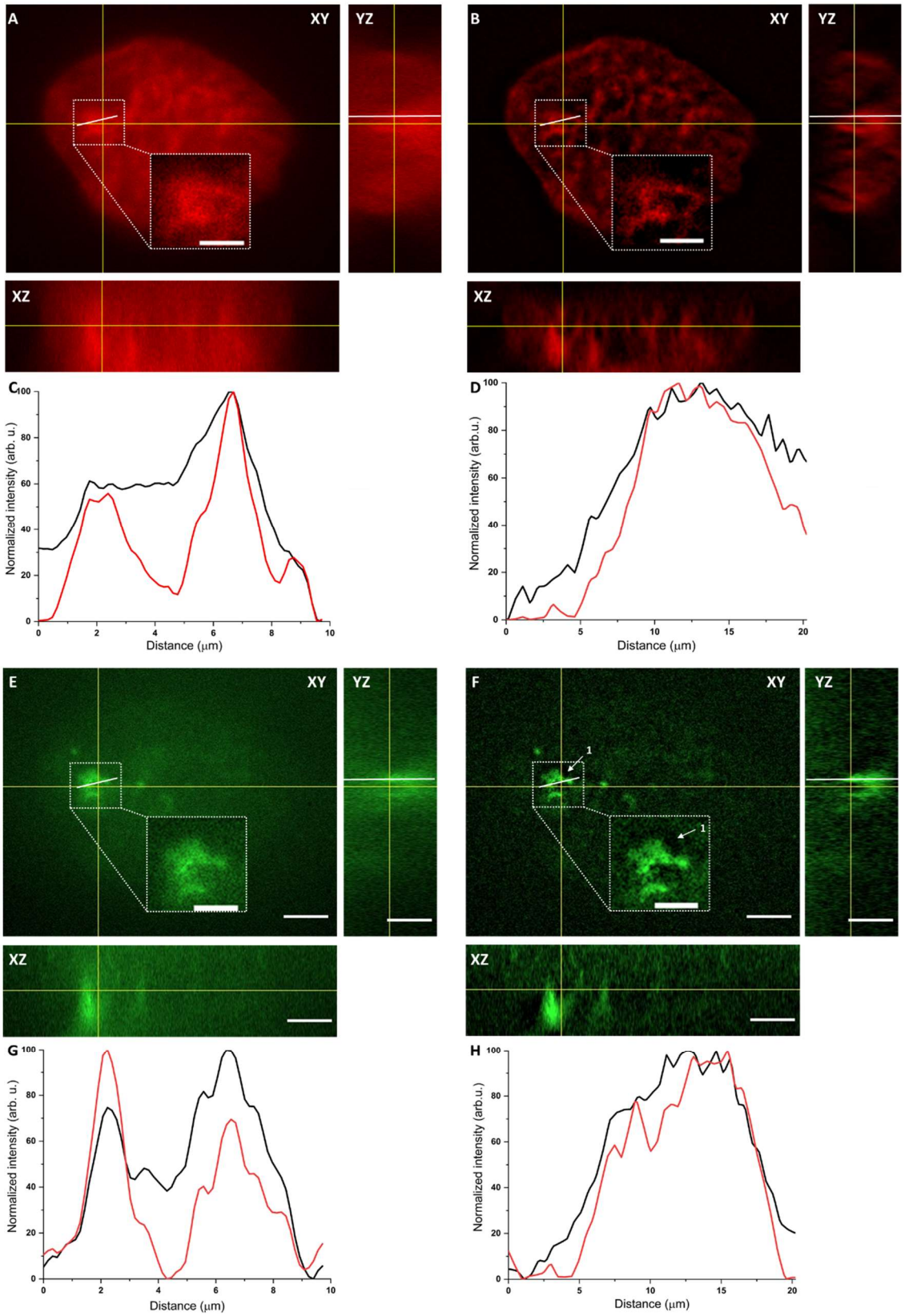


Fig. 71 Comparison of orthogonal planes of M.TaqI labelled MCF10A cell in expanded gel before deconvolution: DAPI (**A**) and M.TaqI (**B**); and after deconvolution: DAPI (**E**) and M.TaqI (**F**). Normalized intensity profiles along the white line in XY plane of DAPI channels (**C**) before deconvolution (black line) and after (red line); in YZ plane of DAPI channels (**D**) before deconvolution (black line) and after (red line); in XY plane of M.TaqI channels (**G**) before deconvolution (black line) and after (red line); in YZ plane of M.TaqI channel (**H**) before deconvolution (black line) and after (red line). Scale bars = 10 μm; in zoom areas = 5 μm, Z step = 0.5 μm.

Deconvolution of the expanded samples improved the resolution and decreased blurring from out-of-focus fluorescence. This, with addition to the expansion microscopy, has provided improved insight into spatial organisation of M.TaqI target sites clusters. However, due to the low axial resolution of the widefield microscopy and mathematical processing of the images which may introduce some artifacts or delete signal we would prefer to observe the spatial organisation of the unmethylated chromatin regions directly. In addition, the processing of a single image is quite slow, especially if more than one channel are used.

Expansion microscopy helped to remove most of the extranuclear background fluorescence in case of M.TaqI labelled cells. The resolution and sharpness of the widefield images were improved after the application of deconvolution algorithm, though with some limitation. Therefore, we tested this approach with M.MpeI labelled cells to hopefully uncover organisation of unmethylated CpG regions of the chromatin.

II.3.2.4. Visualization of unmethylated CpG regions of chromatin using ExM

Due to a significantly higher (10x) cofactor concentration used for labelling with M.MpeI compared to M.TaqI, the background fluorescence intensity (non-specific labelling of nucleophiles in the cell) is significantly higher. Hence, we anticipate that the use of ExM to remove this labelled proteinaceous component of the cell will be even more critical for the CpG-targeting enzyme.

Unmethylated DNA in fixed cells was labelled as previously with M.MpeI MTase and AdoHcy-6-N_3 cofactor (65 μM), followed by conjugation of the Alexa488-alkyne fluorophore using a copper-catalysed click reaction (azide-alkyne cycloaddition). Then,

DNA anchor was introduced and the cells were embedded in the polymer gel. After the digestion step nuclei were stained with DAPI and the gels were expanded in deionized water.

A negative control, where the M.MpeI enzyme was omitted, and the experimental sample were imaged using a widefield microscope before expansion (Fig. 72 A-F). In the images of the negative control sample whole cells can be clearly distinguished in the M.MpeI channel unlike when AdoHcy-6-yne was used as a cofactor at lower concentration (See Fig. 68 A). This suggests that the fluorescent signal from nonspecific labelling with AdoHcy-6-N_3 was not completely removed upon digestion

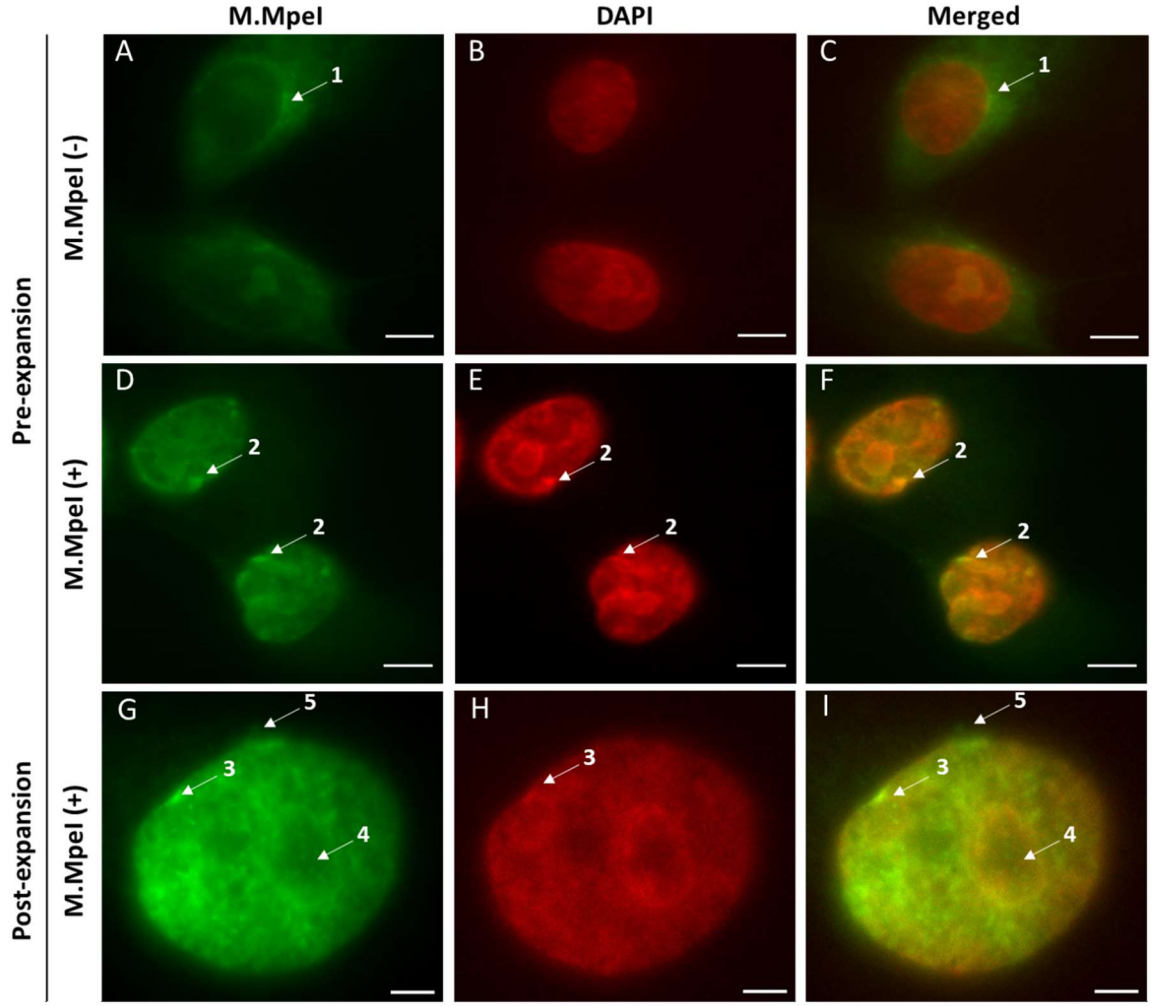


Fig. 72 Widefield images of ExM samples of M.MpeI and AdoHcy-6-N_3 (65 µM) labelled MCF10A cells before (**D-F**) and after expansion (**G-I**). Images of M.MpeI (-) and AdoHcy-6-N_3 (65 µM) samples before expansion (**A-C**). Single slices from Z-stack images are shown. Scale bars = 10 µm.

with Proteinase K. A brighter region surrounding the nucleus is observed (Fig. 72 A,C 1). Although, the origin of this signal is unclear, its intensity is negligible in experimental samples as it is not as bright as the signal originating from M.MpeI labelled DNA (Fig. 72 D,F,G,I).

The structure of unmethylated CG regions of the genomic DNA in the nucleus, compared to M.TaqI labelling, is more dense and more complex. The M.MpeI signal overlaps in many areas with DAPI signal which is reasonable, as higher density DNA regions statistically contain more CpG sites of which a fraction will be unmethylated and labelled. However, some bright regions of labelling, that we hypothesise are parts of the genome rich in unmethylated CpG sites, can be observed (Fig. 72 D,F 2). The image of the expanded sample provides more details on unmethylated CpG organisation. The fluorescence of larger clusters appears coincident with with more intense signal from the DAPI channel implying that these are regions of high DNA density (Fig. 72 G-I 3). Moreover, a multiple small foci can be identified within nucleoli (Fig. 72 G-I 4), which was not observed in M.TaqI labelling. This might be due to a small amount of DNA present in nucleoli, mostly rDNA,[281] where the M.TaqI target sites content was not high enough to give a distinguishable signal.

We applied the deconvolution algorithm to improve the quality of images and obtain more information on the structure of labelled chromatin. Even though a complex organisation was revealed, the signal is blurry and out of the focal plane fluorescence is significant especially in lower fluorescence intensity regions. In addition, some faint objects can be observed outside of the nuclei. However due to a low signal to noise ratio and out-of-focus blur it is difficult to characterise them in more details (Fig. 72 G, I 5).

A single slice from a Z-stack image was compared before and after deconvolution (Fig. 73 A-F). Deconvolved images show decreased out of the focal plane fluorescence and reduced blurring of the visible structures in both channels. Large clusters of M.MpeI labelled DNA, that are colocalised with domains with high intensity emission from DAPI, are now more pronounced with improved signal to noise ratio around these objects (Fig. 73 D-F 1). However, the most dramatic improvement can be observed in the low fluorescence intensity regions. In deconvolved images a signal of unmethylated CpG regions can be observed in the regions of lower DAPI intensity signal – low DNA density regions – which without deconvolution, due to the high noise, were regions with poor contrast (Fig. 73 D-I 2). These regions may have regulatory function, since they have, on average, low DNA density and, hence, lower statistical occurrence of unmethylated CG sequences. Therefore, an intense fluorescence signal would mean that these regions are rich in unmethylated CpG sequences such as CpG islands in regulatory elements e.g. gene promoters. In addition, the fluorescence of the aforementioned single foci in nucleoli are more pronounced (Fig. 73 A-F 3) as well as small clusters outside the nuclei (Fig. 73 G-I 4). These, separated round objects, are mostly noticeable in slices of the bottom part of the cell (closer to where cells were attached to the surface). This is probably due to the shape of the cells grown on the surface. The cells are wider close to the surface of the dish, hence more cytoplasm is

present. Therefore, assuming this signal is not an artifact, it is likely extranuclear DNA fragments e.g. mitochondrial DNA which usually exhibits low levels of CpG methylation.[282]

The deconvolution of the images of M.MpeI labelled cells after expansion suggests a complex structure of CpG unmethylated chromatin. In particular, it helped to identify potentially the high level of CpG unmethylation in low density DNA regions or small object in nucleoli which consist of rDNA or outside the nucleus, which could be possibly

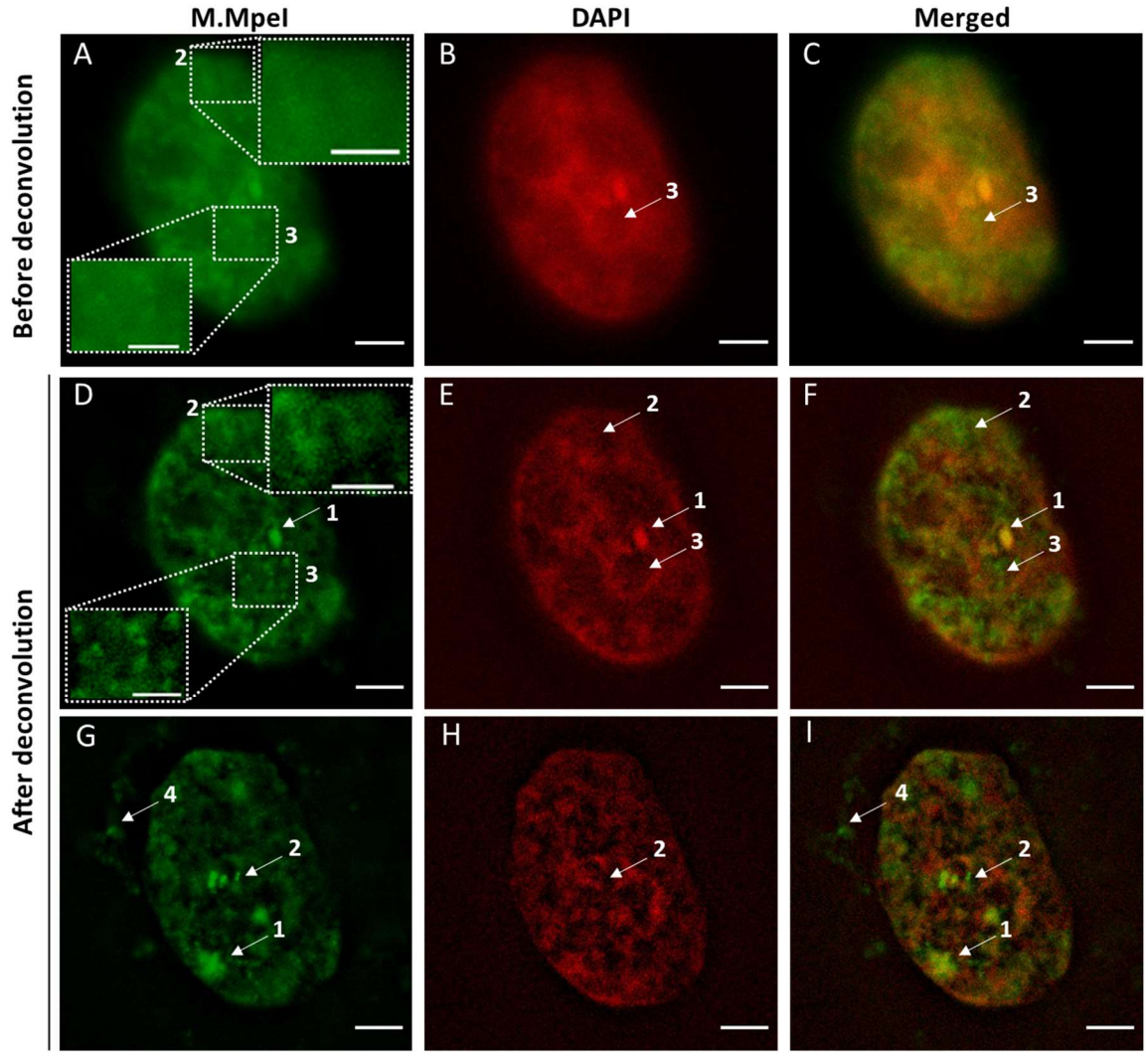


Fig. 73 Widefield images of expanded ExM gel of M.MpeI and AdoHcy-6-N_3 (65 µM) labelled MCF10A cell before (**A-C**) and after deconvolution (**D-I**). Single slices from Z-stack images are shown. **A-F** show the same slice, **G-I** is a slice located towards a bottom of the cell. Scale bars = 10 µm, and 5 µm for zoomed areas.

mitochondrial DNA. However, analysis of the obtained images using a widefield microscope and subsequent processing revealed some limitations.

Firstly, from analysis of the images of M.TaqI labelled cells we showed that the axial resolution is not sufficient for more detailed analysis of the clusters e.g. three-dimensional shape or volume/area due to possible signal bleeding from other slices even after deconvolution. Secondly, visible deconvolution artifacts around bright objects visible as dark envelopes brings uncertainty over possible deletion of the signal. In the case of M.MpeI labelled samples, a variety of structures and object differing in size and shape was revealed with lower intensity objects/patches located in close proximity of the large bright clusters. Therefore, any signal deletion or addition would be more difficult to spot and recognize.

Motivated by these observations we were looking for a technique which enables a fast and direct imaging of three-dimensional structures with no post-acquisition processing needed. To fulfil these requirements we decided to apply iSPIM technique.

II.3.2.5. Application of iSPIM technique for visualization of three-dimensional architecture of unmethylated M.TaqI target sites of chromatin

Light sheet microscopy, such as SPIM, is known to greatly reduce noise from out of the focal planes and enables fast, volumetric imaging. Most importantly, the relatively low photon dose per unit volume minimised photobleaching of the sample, relative to

confocal imaging. We applied the iSPIM technique, in combination with ExM, to directly visualize the three-dimensional organisation of unmethylated chromatin regions in the nuclei of fixed cells.

As previously, we tested the technique on MCF10A cells labelled with M.TaqI and AdoHcy-6-yne cofactor. As a result of imaging using iSPIM, a stack of single sheet images is obtained, from which a three-dimensional image can be reconstructed. Firstly, an image of a negative control, M.TaqI (-), where cells were incubated with AdoHcy-6-yne cofactor in absence of M.TaqI enzyme, was compared to an experimental sample where M.TaqI enzyme was added (Fig. 74 A-F). Noticeably, a homogeneous bright signal from a nucleolus is present in the M.TaqI channel of the negative control, which is in accordance with previously recorded images of this sample (Fig. 74 A, C 1). However, in contrast to those images recorded using the widefield setup (Fig. 68), image contrast is greatly improves, with isolated small clusters of labelling inside and outside of the nucleus now being visible (Fig. 74 A,C 2). A reduction of out-of-focus signal is distinct in DAPI channel, where a single sheet image resembles the deconvolved widefield images. In the M.TaqI labelled cell, relatively bright fluorescence, likely from encapsulated fluorophores in a nucleolus can also be observed (Fig. 74 D, F 1). Unlike in M.TaqI negative control, small isolated

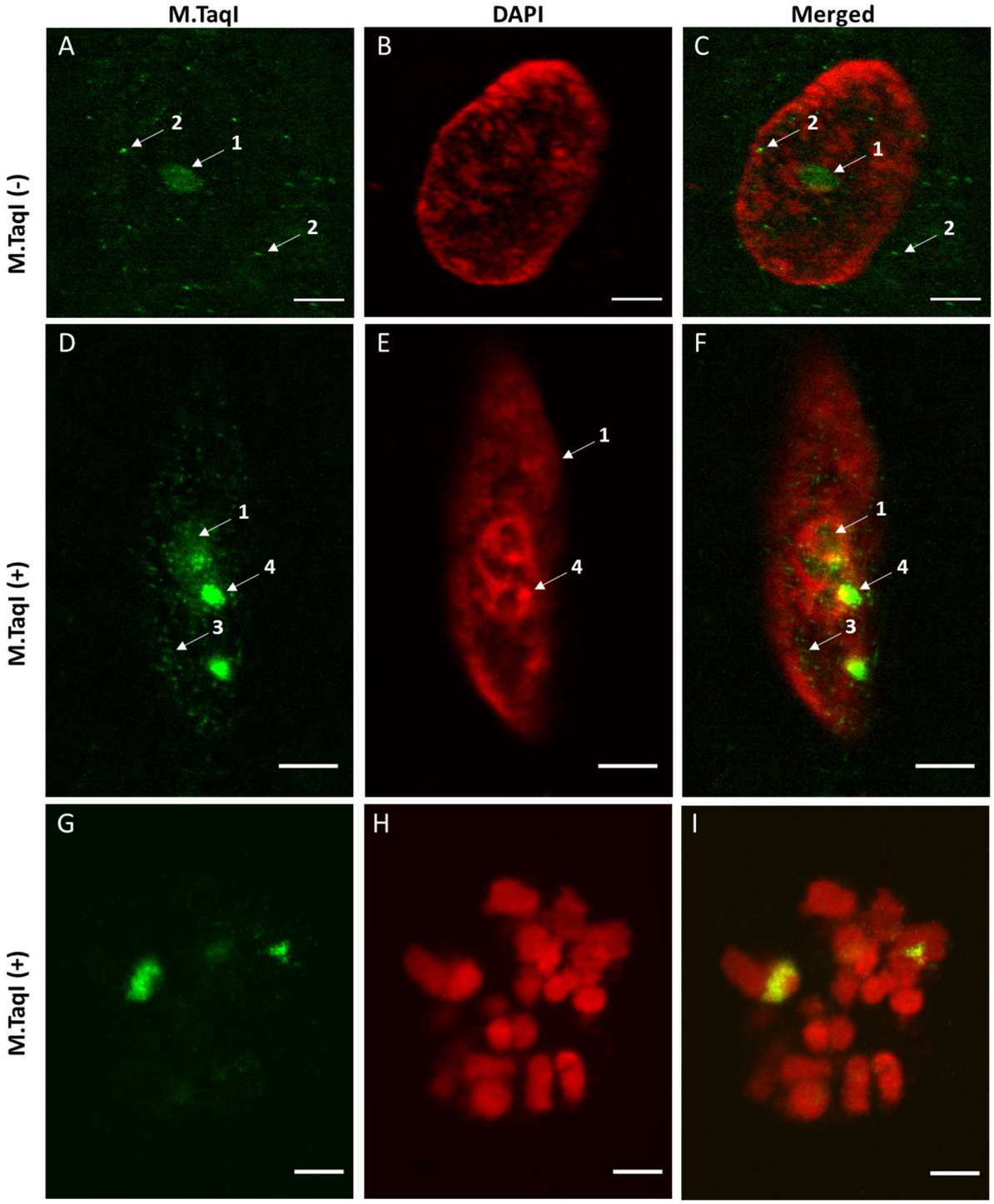


Fig. 74 iSPIM images of expanded gel samples of MCF10A cells: M.TaqI (-) control (**A-C**) and M.TaqI (+) (**D-I**) AdoHcy-6-yne (3 μM). Single planes of stack images are shown. Scale bars = 10 μm.

clusters occur only within the nucleus (Fig. 74 D, F 3). This suggests that they are M.TaqI-labelled regions of the genome that have relatively low labelling site density. These can be resolved, due to the improved signal-to-noise ratio in the iSPIM images.

In addition, large clusters, colocalized with dense regions of DNA, are also visible as in the widefield images (Fig. 74 D-F 4). The range of signal intensities in these images, was adjusted to show low intensity objects and, hence, the signal from densely-labelled regions of the nucleus appears saturated in the Fig. 74 D, F, G, I.

An image of a mitotic cell with forming chromosomes was captured (Fig. 74 G-I). This is important for several reasons. Firstly, DNA in mitotic cells is organised in dense chromosomes with limited accessibility. Such an image could provide information on the efficiency of the labelling of dense regions of DNA in interphase cells. Secondly, a mitotic cell has two structurally-similar copies of the chromosomes. Therefore, if M.TaqI labelling is specific and robust we should be able to observe the same spatial labelling distribution on both copies of a given chromosome. Although the structures visible in the DAPI channel appear to be partially-condensed chromatin and, hence may not represent fully segregated chromosomes, for simplicity we will refer to them as chromosomes when describing their organisation or structures.

It can be observed that M.TaqI is able to recognise, access and label DNA in a condensed chromatin (Fig. 74 G-I). A large cluster of labels is visible on what appears to be a single chromosome. This raises a question on a definition of ‘accessible’ genome. One explanation would be that the labelled area is not fully condensed or this regions is somehow exposed to the labelling. However, the accessibility of the genome to the labelling with M.TaqI could relate to the primary structure of chromatin (‘beads-on-string’), in which a tightly wrapped DNA around histone core is not accessible to the enzymes and therefore cannot be modified. The global condensation of chromatin however, is not homogenous and fluctuates between various levels of condensation.[283]

It was shown that condensed chromatin is still accessible to proteins associated with nucleosomes such as transcription factors.[284]

The symmetry in the spatial distribution of labelling that we predicted is clearly evident in the three-dimensional projection of the mitotic cell (Fig. 75). This image was rendered in ChimeraX software to better show the structure of such compacted chromatin. The M.TaqI labels are indeed located on a single chromosome (Fig. 75 A 2). Furthermore, this cluster of labels seems to be located close to the central part of this chromosome as the chromosome arms can be recognized (Fig. 75 A 1). Hence, we would predict that, two copies of the same DNA sequence should be present, with one on each arm of the chromosome. The symmetrical and evenly distributed label density on this part of the chromosome is consistent with effective and specific M.TaqI-directed labelling. The symmetry of the labelling can also be noticed in other part of the nucleus (Fig. 75 B). This region displays potential symmetry across the dotted blue axis with two characteristic arms in the DAPI channel on the opposite sites (Fig. 75 B

3a and 3b). The M.TaqI labelling within this region shows clear symmetry with a narrowing at the axis.

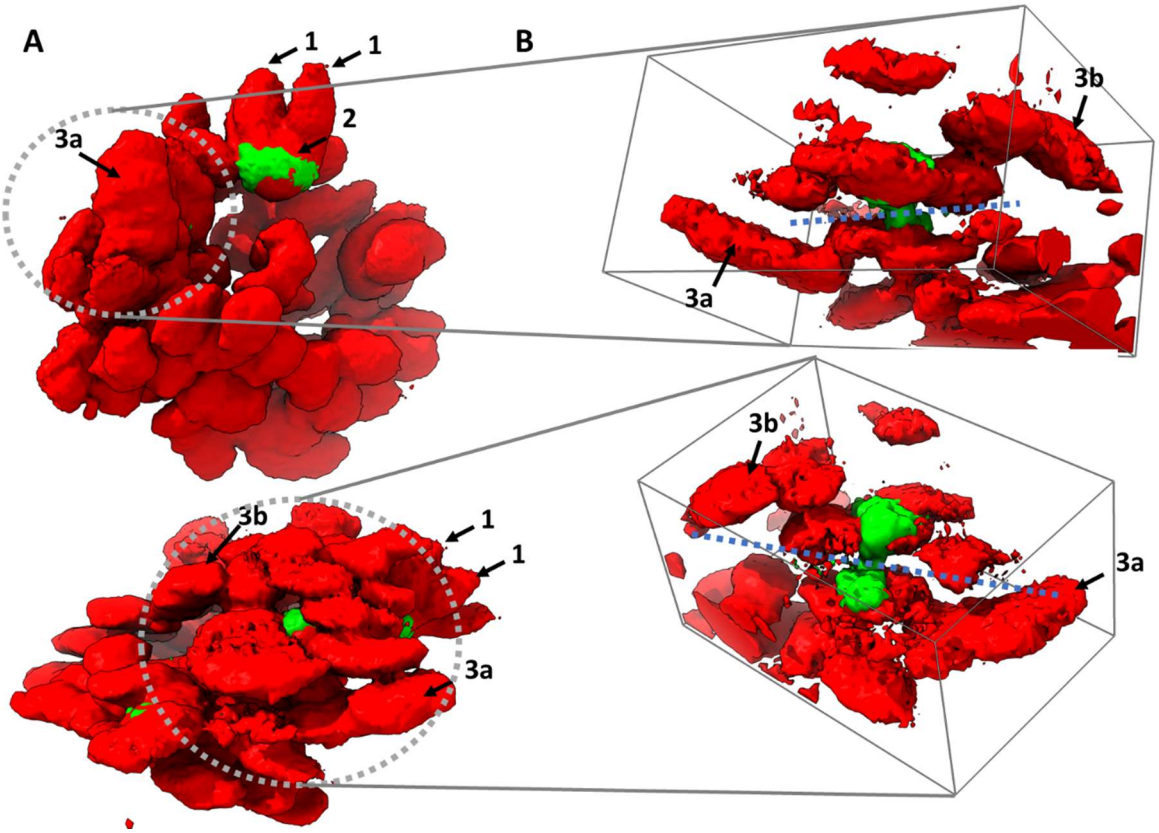


Fig. 75 3D rendered iSPIM image of expanded MCF10A cell with condensed chromatin labelled with M.TaqI and AdoHcy-6-yne (3 μM): (**A**) volumetric image of the nucleus; (**B**) a cropped volume; red – DAPI, green – M.TaqI. ChimeraX was used for rendering

Three-dimensional projections have been reconstructed from the recorded single lightsheet images (Fig. 76). In addition to the clusters with complex architecture and colocalized with nucleoli, in these examples (Fig. 76 A, D 1), smaller objects are evenly distributed across nuclei and are distinguishable as a result of a high signal to noise ratio in the images (Fig. 76 A, C, D, F 3). As previously observed in the single plane images, these features are exclusively contained within the volume of the nucleus based on DAPI staining. This is consistent with our expectations, that M.TaqI labelling sites should be present across the genome at a relatively consistent density, outside

of the ‘hotspots’ of labelling sites that occur in some subset of repetitive regions of DNA sequence.

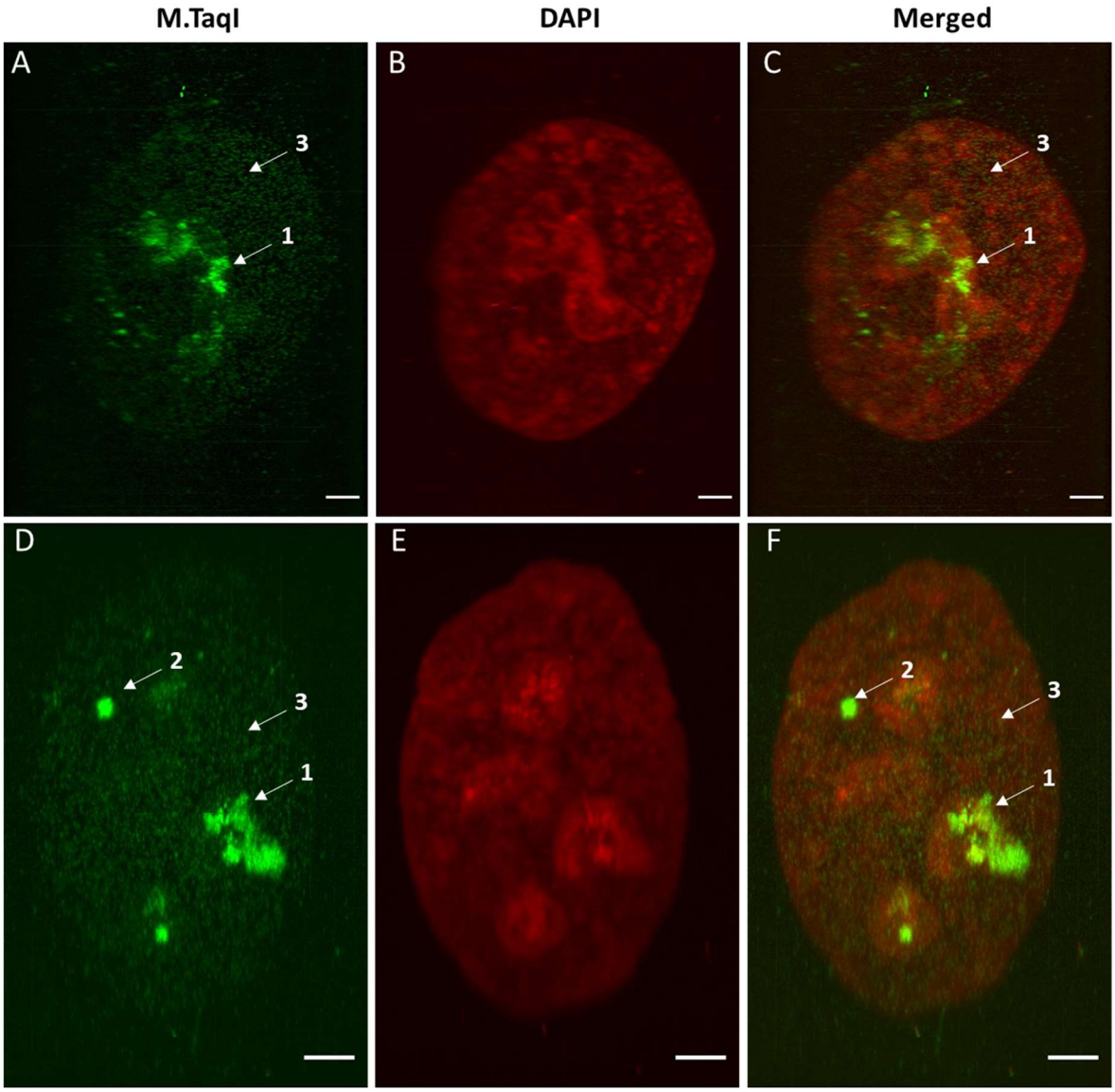


Fig. 76 3D projections of iSPIM images of expanded MCF10A cells labelled with M.TaqI and AdoHcy-6-yne (3 μM). Scale bars = 10 μm.

These projections show also some difficulties of using iSPIM. In both images, especially in DAPI channel, ends of the nuclei (left side of the top image and bottom of the bottom image) are slightly out of focus. This can be caused by sample drift during measurement or calibration.

II.3.2.6. Application of iSPIM technique for visualization of three-dimensional architecture of unmethylated M.MpeI target sites of chromatin in MCF10A and MCF7 cells

We also used iSPIM to image MCF10A and MCF7 cells labelled with M.MpeI, with an unmethylated CpG dinucleotide sequence as a target, and AdoHcy-6-N_3 as a cofactor. In the negative control, M.MpeI (-), where M.MpeI was omitted from the labelling reaction and cells incubated with AdoHcy-6-N_3, fluorophores within nucleoli and an evenly distributed background labelling across the cell are observed (Fig. 77 A-C 1, 2). However, in M.MpeI (+) samples of both MCF10A and MCF7 cells, where samples were incubated in presence of M.MpeI and AdoHcy-6-N_3, a signal to noise ratio is high enough so the background signal is negligible. Improved contrast, in comparison to the widefield images, helps to confirm the presence of multiple clusters of M.MpeI target sites and also provides more details on the structure of unmethylated CpG in the genome without the need for post-processing. Labelling clusters can be directly observed in the nucleoli as well as multiple small objects outside the nucleus (Fig. 77 D, F, G, I - 3, 4) in contrast to the widefield images (Fig. 73 D, G). The iSPIM images provide a further insight into the organisation of unmethylated CpG regions of chromatin. In MC10A cells, within the complex network-like structure they seem to be organized in smaller globular domains of various size (Fig. 77 D, G 5) from ~0.5 μm in diameter for small foci up to large clusters of ~6 μm in diameter (Fig. 78 E) in expanded samples. By contrast, in cancer cells (MCF7), a high intensity contrast is observed between large clusters, a few μm in diameter (Fig. 77 G, I 5), and foci within a nucleolus

(Fig. 77 G, I 4) which tend to be brighter than the rest of the nucleus. Furthermore, outside of a nucleus, in addition to small clusters of labels (Fig. 77 G, I 4), significantly

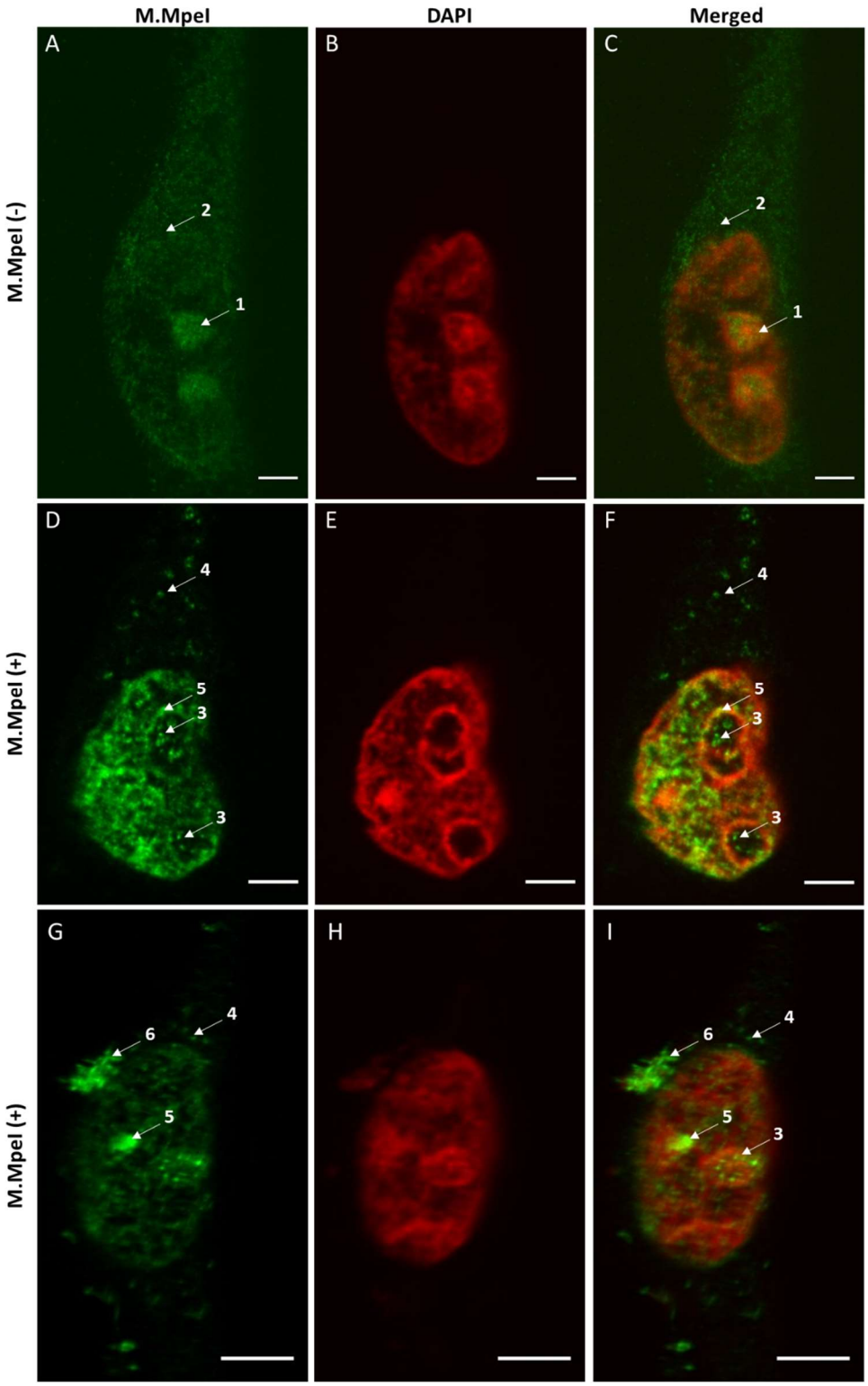


Fig. 77 iSPIM images of expanded gel samples of MCF10A cells: M.MpeI (-) control (**A-C**) and M.MpeI (+) (**D-F**) and MCF7 cells : M.MpeI (+) (**G-I**); AdoHcy-6-N_3 (65 µM). Single planes of stack images are shown. Scale bars = 10 µm. M.MpeI channel was linearly translated to match DAPI channel.

larger clusters of labels, a few µm in diameter, are located (Fig. 77 G, I 6).

Various parts of the cell can be analysed separately when analysing single slices of iSPIM images, providing some insight into the small, bright labelling clusters in nucleoli. However, a more general view on the organisation of M.MpeI target sites is provided by a three-dimensional projection.

A presentation of a three-dimensional projection as a heatmap of the signal intensity provides intricate details on both characterisation of the spatial organisation of the globular domains and a relative density of labelled CpG sequences in MCF10A cells (Fig. 78). The intensity range was adjusted separately for each image in order to show the difference in relative signal intensity within each cell. In the case of presented cells, unmethylated CpG fragments of the chromatin form several bright, dense labelling clusters per cell (Fig. 78 red) with the large regions of medium intensity (fewer unmethylated CpG sites) around them (Fig. 78 yellow/green). In a few of the presented examples (Fig. 78 A-E) labelled regions appear to be pulled towards the poles of the nuclei. In some of the presented nuclei, regions of higher density of M.MpeI target sites are located on one side of the nucleus with red clusters shifted towards periphery of the nuclei (Fig. 78 C, E). In one of the images, a nucleus lacks large, bright clusters, and domains of 1-2 µm diameter and medium intensity (yellow/green) distributed more evenly across the nucleus can be found (Fig. 78 F). A size, number and distribution of both large clusters and smaller domains might depend on the dynamic changes in the activity of transcription machinery depending on the cell cycle e.g. high transcription rates in the growth phase (G_1) or low transcription in DNA synthesis phase (S). Although all shown cells in Fig. 78 are in the interphase, the variety of the structures found can be an effect of cells being at different stages of the interphase. Due to the

scarcity of data presented, the observation of size and distribution of labelled chromatin regions cannot be generalised to the whole population of MCF10A cells.

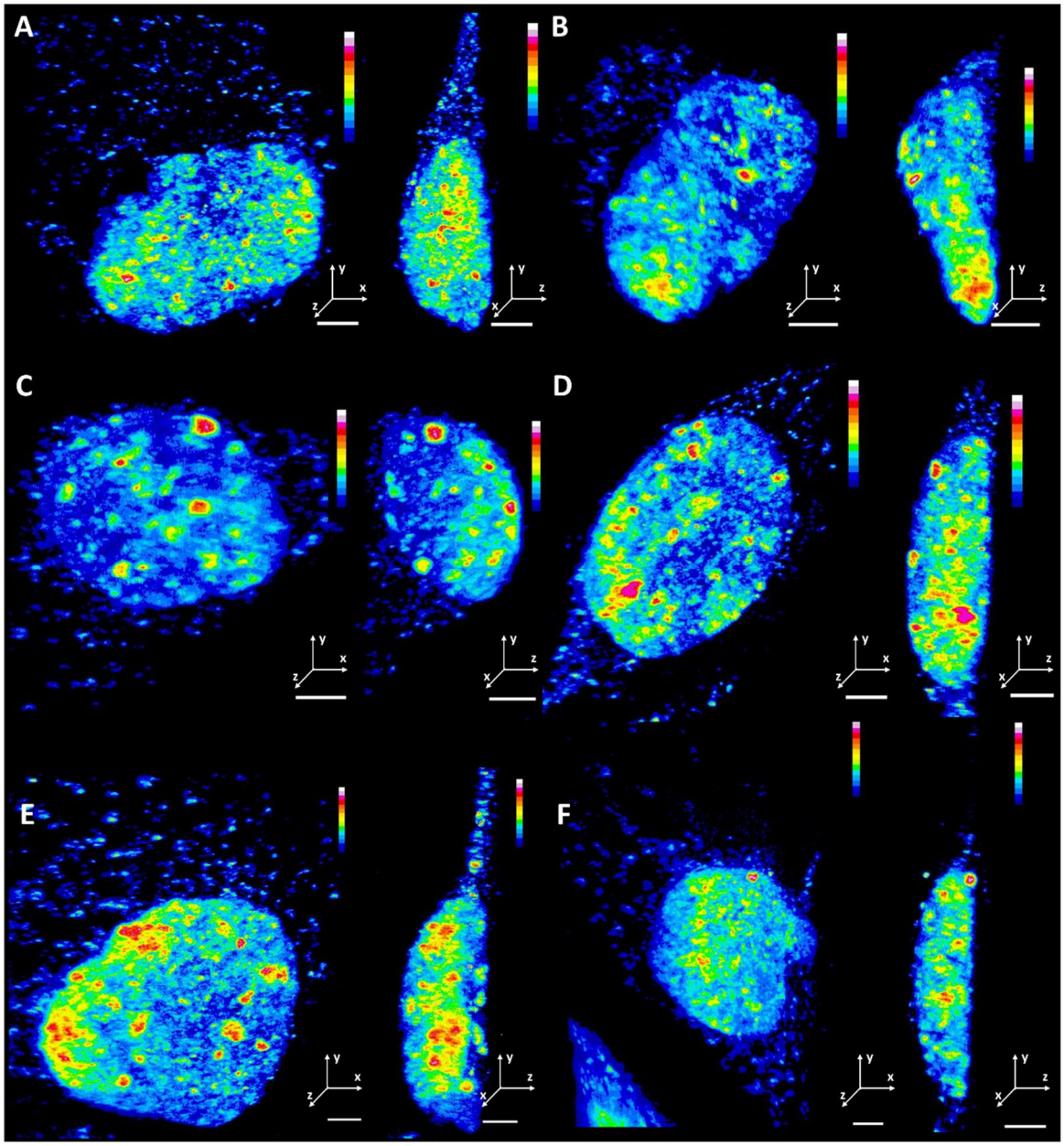


Fig. 78 3D heatmap projection of iSPIM images of expanded MCF10A cells labelled with M.MpeI and AdoHcy-6-N_3 (65 µM). a view from the top of the cell (left panels a side view (right panels). Scale bars = 10 µm.

An interesting image of two daughter cells in telophase was captured (Fig. 79). The image shows remarkable symmetry of the labelled regions with some clusters of labels present in the same area of the opposite cells (Fig. 79 A 1). This is good evidence to

support claims that the labelling by the M.MpeI enzyme is specific. Regions containing high densities of unmethylated CpG sequences are localized in the further ends of each cell even though, the distribution of the densely packed chromatin is quite even across both nuclei (DAPI channel).

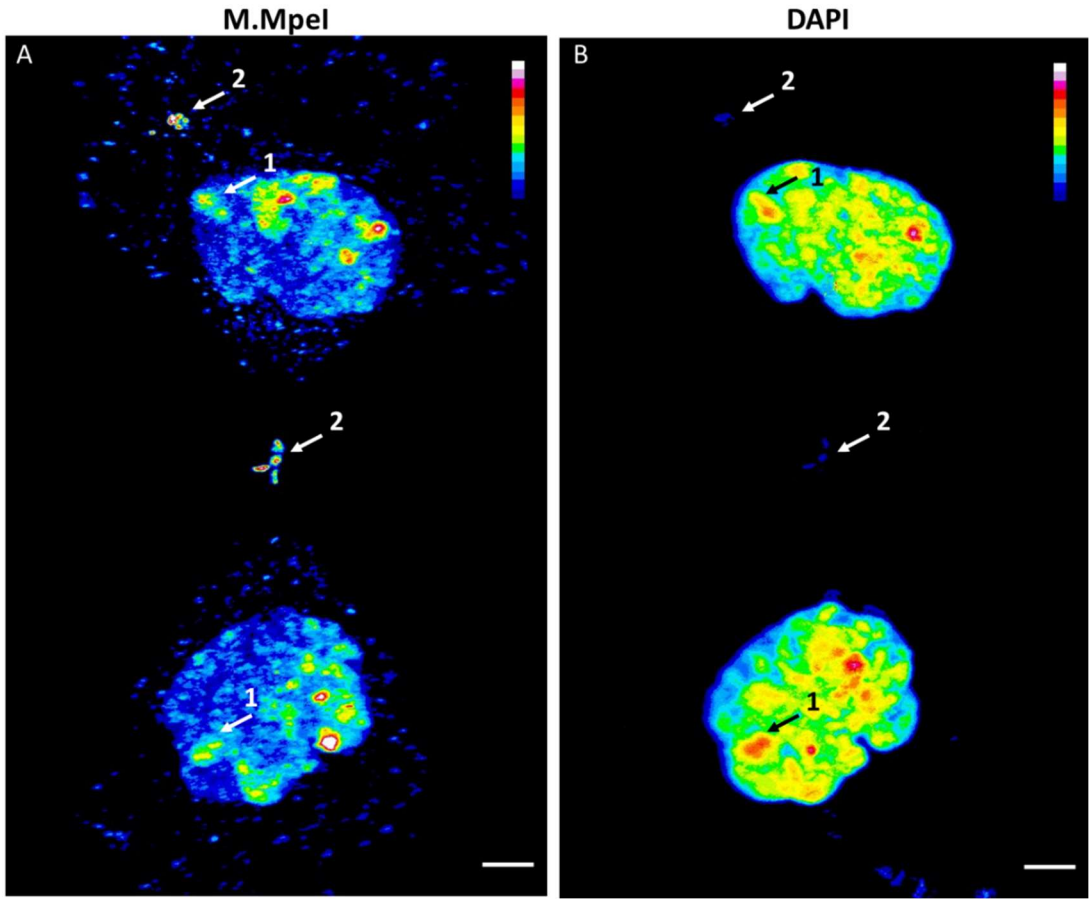


Fig. 79 3D projection of iSPIM images of expanded MCF10A cell labelled with M.MpeI and AdoHcy-6-N_3 (65 μM). Scale bars = 10 μm.

What attracted our attention were two, a few μm in size, clusters localized out of both nuclei, visible in both M.MpeI and DAPI channels (Fig. 79 A, B 2). In the literature there are reports of so called extrachromosomal circular DNA (eccDNA). The eccDNA sequences ranging in size from 10 bp up to 20 kbp, were identified in plants and animals including human.[285] Numerous of characteristic sequences have been found in eccDNA including gene coding sequences i.e. rDNA and histone genes, satellite repeats,[286] 5' untranslated regions of genes or CpG islands.[287] They are thought to be

responsible for genome plasticity and play a role in the evolution of the genome. A greater number of eccDNA, encoding oncogene sequences, have been found in various cancer cells. Although still little is known about its function in cancer, recent studies suggest that eccDNA, as a carrier of oncogenes, can be extensively amplified in cancer cells leading to the overexpression of the oncogenes and tumour progression.[288] Therefore, the ability of the developed method to detect unmethylated eccDNA could find diagnostic application. However, with this approach it cannot be concluded that observed clusters are fragments of eccDNA.

In shown 3D projection images of M.MpeI labelled cancer cells (MCF7) a number of labelled clusters of unmethylated DNA, of several microns in size, containing unmethylated CpG sites outside the nucleus. An extreme example is shown in Fig. 80 F where at least eight clusters with diameters up to 4 µm of extranuclear DNA are present. In addition, in all the examples, such clusters are characterised by a high relative fluorescence intensity (red) suggesting the high density of unmethylated CpG sites and accessibility to labelling by M.MpeI (Fig. 80 A-F). Hence, their methylation state and accessibility likely promotes high expression efficiency, which is desirable for rapid tumour progression. This is in concordance with a high number of copies of eccDNA in cancer cells, however more evidence is needed to confirm that these objects are molecules of eccDNA.

The presented images of MCF7 cells, DNA containing M.MpeI target sites is localised in clusters of a high relative fluorescence intensity (red) whereas the rest of the nucleus shows weak relative fluorescence (blue/green) (Fig. 80 A-E). This is surprising as DNA of cancer cells is considered to be hypomethylated on a genome-wide scale. However, when a MCF7 and MCF10A cells are presented using the same intensity range, the

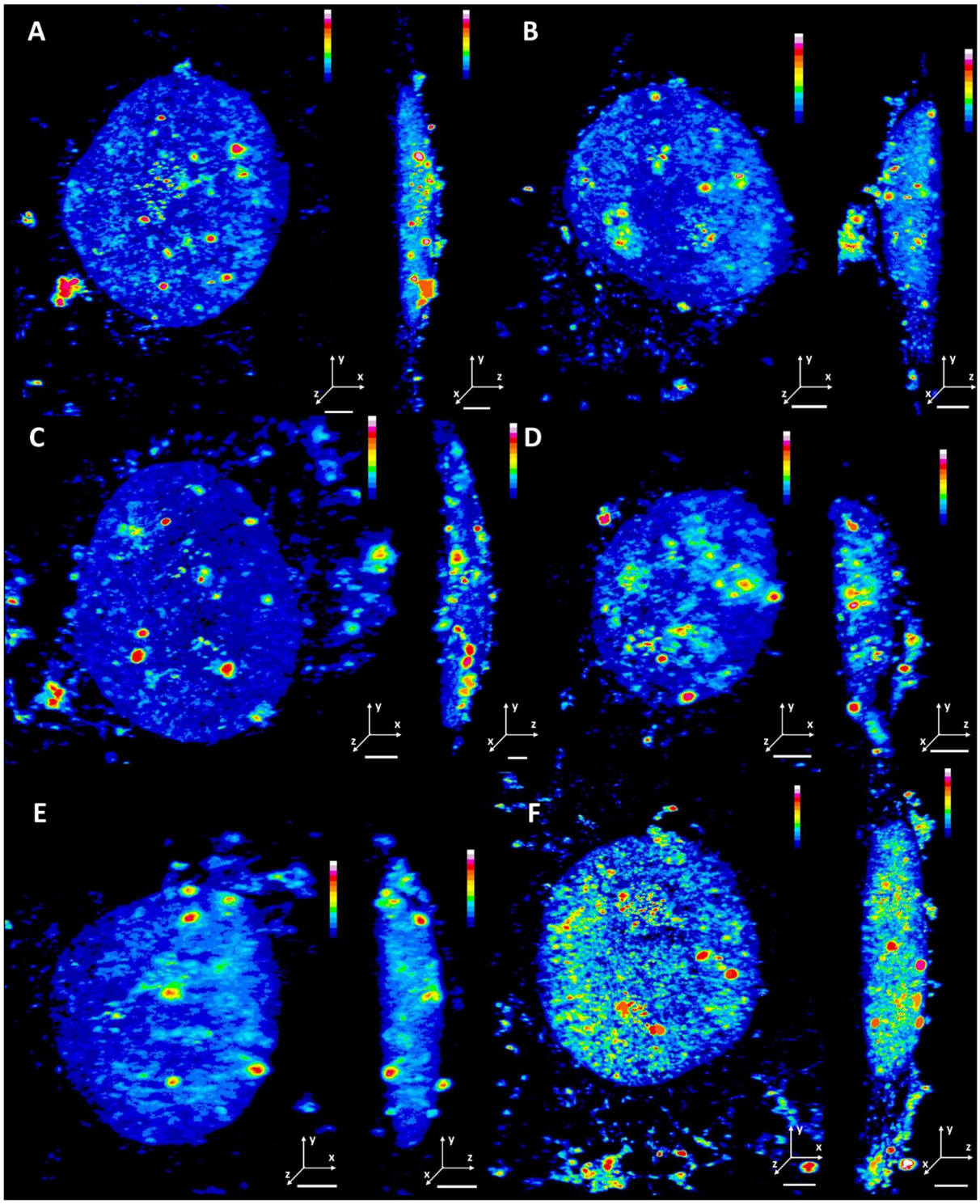


Fig. 80 3D heatmap projection of iSPIM images of expanded MCF7 cells labelled with M.MpeI and AdoHcy-6-N_3 (65 µM). a view from the top of the cell (left panels a side view (right panels). Scale bars = 10 µm.

regions of lower density of M.MpeI target sited are comparable in intensity (Fig. 81 green/yellow) whereas the greatest difference is observed in the clusters of the highest density of M.MpeI target sites. These in MCF7 show intensity of 154 grey value (Fig.

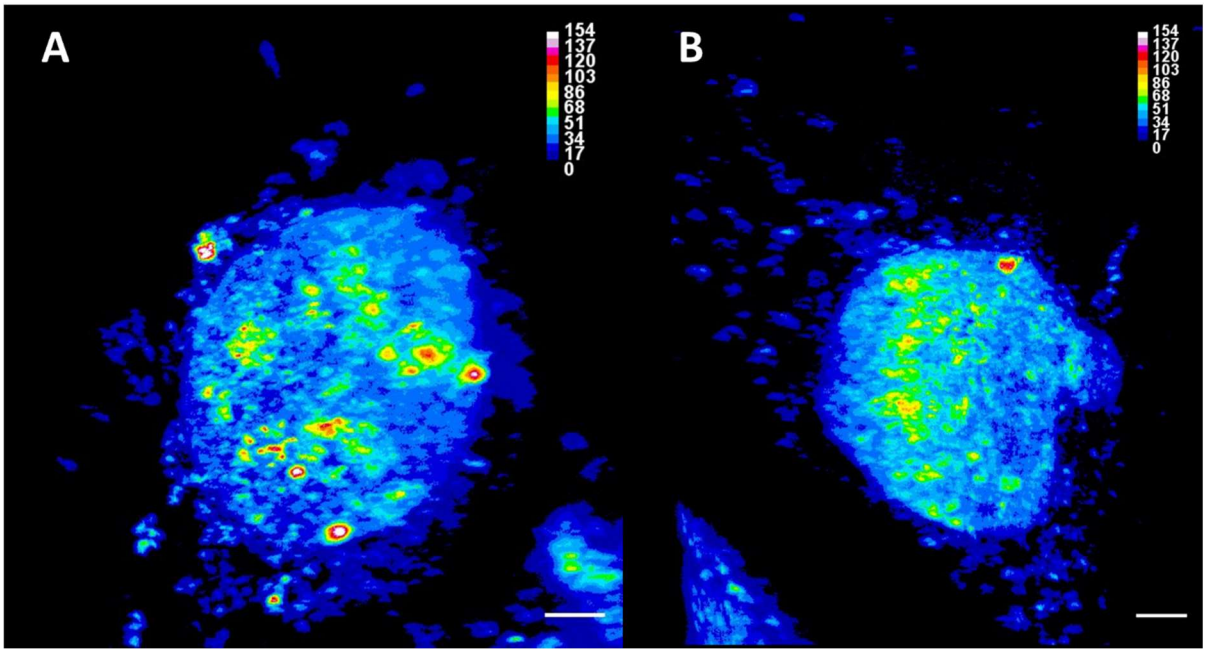


Fig. 81 3D heatmap projection of iSPIM images of expanded MCF7 cell (**A**) and MCF10A (**B**) labelled with M.MpeI and AdoHcy-6-N3 (65 μM). The range of the signal intensity is the same on both images. Scale bars = 10 μm.

81 A - white) whereas in MCF10A is about 22% lower, at 120 grey value (Fig. 81 B-red). The images of these cells were taken separately using the same microscope setup but without an internal intensity standard hence, such comparison is only an estimation. It could suggest that the labelling density of such clusters increases, likely due to decreased methylation, in breast cancer cells (MCF7) compared to normal counterparts (MCF10A). However, in order to support this claim, a large set of images need to be taken and statistically analysed in the future. These highly unmethylated clusters may contain important genes for cancer progression which require intensive expression (are overexpressed in these cells) similarly to very active and hence, bright clusters of unmethylated DNA outside the nucleus, which could be eccDNA. In shown examples, the distribution of labelled clusters within the nucleus varies with some being located closer to the periphery of the nuclei whereas others are found more in the internal part (Fig. 80 A-F). It is commonly agreed that periphery of the nucleus is enriched in densely-packed and less active, hence methylated chromatin, whereas an

active and open chromatin is localised in the interior of the nucleus. However, there are examples of gene activation in specific loci inducing the repositioning of this area of chromatin to the periphery of the nucleus within the chromosome domain.[289]

The shown images show a variety of distribution patterns across the nuclei in both MCF10A and MCF7 cells. A common feature is a presence of globular like structure with high intensity cluster of size up to a few micrometres. Based solely on the presented images, in MCF7 cells, unlike in presented images of MCF10A, large extranuclear cluster of high intensity can be observed. However, due to the lack of statistical analysis, the organisation of the M.MpeI labelled DNA cannot be compared between both cell lines at the population level.

II.3.2.7. A note on the isotropy of expansion

A common question when applying ExM is the isotropy of the expansion of structures in the imaged specimen. Ideally, gel matrix and embedded structures will expand equally in all three dimensions. This is especially important when absolute measurements of imaged structures are required. Distortions to the specimen structure can be introduced by an inhomogeneous polymer network, incomplete or uneven homogenisation of the sample or an improper mechanical handling of gel samples.[290] There have been a multiple studies where distortions in expanded samples were identified and measured. [272, 273, 275] These distortions occur depending on the variant of ExM method applied, the fixation method and type of biological samples. Anisotropy of expansion is typically small, for example, for a standard gel formulation, introduced by Boyden and co-workers,[270] the distortions of spatial organisation are 5-10 nm on the micrometer scale.[291]

Such measurements are commonly taken via a correlative imaging where cellular structures of the same cells are imaged and compared pre- and post-expansion. For this high resolution imaging techniques are used, in order to enable comparison of changes to imaged structure on the nanometre scale.

In order to identify deformations introduced in the structures of MTase labelled DNA we attempted a correlative imaging using a combination of widefield microscopy and deconvolution. The first challenge was identifying the location of the same cell pre- and post-expansion in a transparent gel. To tackle this, we cut a gel in a specific shape such as a triangle, and imaged cells of a characteristic shape located close to a corner of the gel. An effective field of view is more limited in post-expanded samples e.g. only part of a cell is seen at once, hence orientation and navigation through the sample is challenging. This complicates the localisation of the target cell, thus we decided to first take images in post-expanded sample, collapse a gel in 10x PBS buffer and image the same cells in the shrunken gel. In this way, conservation of the structure during the expansion cannot be completely assessed, as no image is taken prior to expansion occuring. However, we do measure the impact of the gel swelling on the structure of the nuclei in the gel. Furthermore, literature examples as well as the results in our group suggest that observed distortions are usually lower than 1% on a micrometre scale.

For this we labelled MCF10A cells with M.MpeI and AdoHcy-6-N_3 and stained DNA with DAPI. Due to the fact that DAPI is not covalently bound to DNA strands but rather binds to the minor groove, its affinity for DNA may vary during the expansion or the shrinking step leading to inconsistent staining. Hence, the main focus was put on M.MpeI channel.

Correlative imaging shows that the structure of the nucleus is maintained in both expanded and collapsed gels (Fig. 82). The only major differences between images from the M.MpeI channel in both cells (Fig. 82 A, C) are a translational mismatch of the same structure (Fig. 82 A 1) or the lack of signal in either the collapsed or expanded sample (Fig. 82 A, C 2). Both are most likely a result of the precision of the matching of imaging planes rather than it being caused by major distortions of the structures in the nuclei. As images of cells in both states of gel were taken independently it is not possible to image exactly the same focal plane with a reasonable Z-step during acquisition. Furthermore the depth of field in expanded and collapsed samples is different. A fluorescence signal form the DAPI channel of the images in the collapsed gel is, as expected, lacking in many areas of the imaged nuclei. Again, the major differences are translations (up to 1 µm) or a loss of a signal. Due to the relatively dense labelling of the DNA with DAPI, as compared to introduction of a fluorescent tag

in unmethylated CpG sequence, the mismatch between focal planes is even more important in this channel, which explains the local shifts.

We decided these results provide enough evidence to assume that any distortions will not have a significant impact on the analysis of the structures of unmethylated DNA regions in a later part of this chapter. However, the analysis of the expanded versus the collapsed gel does not provide information on whether the densely labelled unmethylated DNA regions expand in the same way as the rest of the cell.

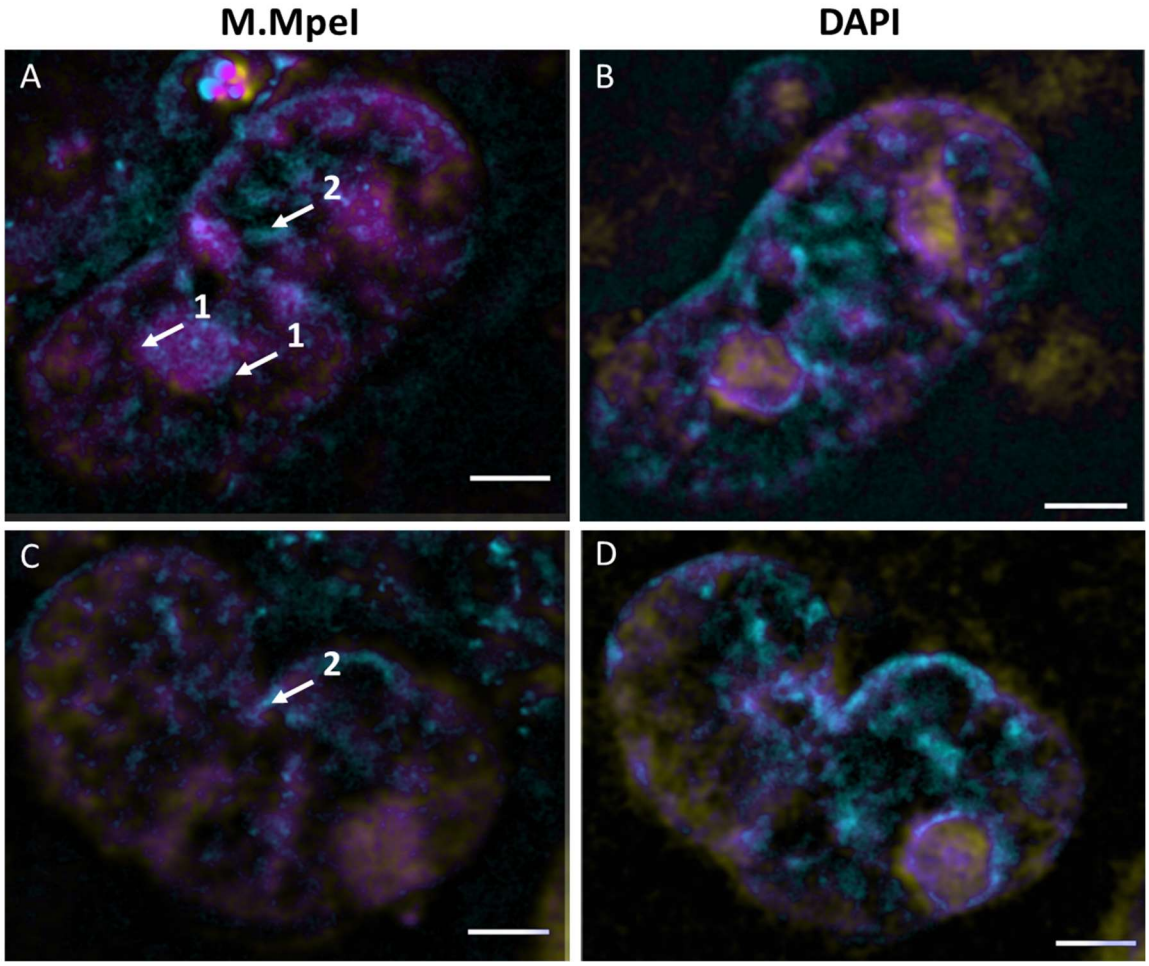


Fig. 82 Correlative imaging of two cells in a shrunk gel (yellow) and post-expansion (cyan). A difference between both images are shown as magenta. Scale bars = 10 µm. Single Z-stack planes are showed.

II.3.3. Conclusions

In order to tackle a high non-specific background signal and get an insight into structure of unmethylated DNA in cells we applied expansion microscopy method with ~4x expansion factor. Expansion microscopy removed proteins from the cells that were labelled non-specifically with our AdoMet analogues and therefore allowed us to resolve the underlying epigenome. Deconvolution was also applied to further improve the resolution of the images.

This method was also applied to M.MpeI labelled MCF10A cells which revealed a complex structure of possibly unmethylated CpG regions of DNA. In the nucleus, dense regions of globular clusters of labelling sites were identified. In addition, fluorescent foci in nucleolar regions were observed which we suggest is likely unmethylated ribosomal DNA. Moreover, fluorescent clusters of DNA were identified in the extranuclear regions which could consist of mitochondrial DNA.

In order to be able to directly image the three-dimensional structure of unmethylated DNA, with no need of post-processing we applied light sheet microscopy (iSPIM). An image of a MCF10A cell with condensed chromatin showed that M.TaqI labelling is effective even in condensed parts of the chromatin.

We applied light sheet microscopy to acquire a first three-dimensional visualisation of accessible unmethylated CpG regions of DNA in human cells. iSPIM images of M.MpeI labelled MCF10A and MCF7 cells provided insight into the three-dimensional structure of probably unmethylated CpG regions. A variety of distribution of globular structures of diameter in arrange of ~0.5 – 6 μm was identified in both cell lines. In MCF10A cells, heavily labelled bright clusters were identified with intensity of the signal gradually

decreasing when moving away from the brightest clusters. in the internal parts of the nucleus. On the contrary, in cancer cells an increased intensity of the brightest clusters compared to the healthy cells was observed. In addition, an increased number of large clusters of labels, up to 4 μm in size, highlight the presence of unmethylated extranuclear DNA in the shown images of cancer MCF7 cells. However, due to the scarcity of data, more experiments are needed to obtain a statistically valid model of organisation of observed domain on the population level. We suggest this is might extrachromosomal circular DNA which is responsible of overexpression of oncogenes and tumour progression, although this needs to be investigated further.

II.3.4. Materials and methods

iSPIM microscopy

iSPIM imaging of expansion gels was performed on an ASI RAMM microscope with SPIM-MOUNT, equipped with Nikon 40x objectives attached to piezo object movers to scan the sample. Illumination was from a 100mW OBIS CW lasers. A Hamamatsu Flash 4 scientific CMOS camera was used for imaging. Micromanager was used to control the system and scan the sample.

Synthesis of LabelX

LabelX was synthesized according to the method from Chen et al., 2016.[273] Briefly, 25 μL of 10 mg/mL solution of 6-((acryloyl)amino)hexanoic acid, Succinimidyl ester (AcX; ThermoFisher) in dry DMSO was mixed with 2.5 μL of 1 mg/mL solution of *Label* IT® Amine Nucleic Acid Modifying Reagent (Mirus Bio) in Reconstitution Solution and incubated overnight at room temperature with shaking. The resulting LabelX solution was stored at -20ºC.

DNA alkylation in cells using bacterial methyltransferase enzymes

Cells were grown and labelled as previously described in II.2.5.

For widefield microscopy on the slides cells were labelled as above and stained with 2 μM DAPI in PBS for 5 minutes in dark and mounted onto glass slide with ProLong™ Gold Antifade mountant.

For ExM cells were incubated with 0.006 mg/mL LabelX in 20 mM MOPS buffer pH 7.5 at 37°C overnight, followed by washing with PBS. The gelation protocol was adapted from Chen et al.[270] Glass cover slip with cells was placed on top of 90 μL of cold and well mixed polymerization solution (1x PBS, 0.2 M NaCl, 8.6% (w/w) sodium acrylate, 2.5% (w/w) acrylamide and 0.15% (w/w) N,N′-methylenebisacrylamide), 0.2% TEMED (w/w) and 0.2% APS (w/w) added last to the solution) for 10 minutes at room temperature followed by incubation at 37°C for 2 hours under humidified conditions. A formed gel with cover slip was then transferred to 2 mL of proteinase K solution (8 U/mL) in digestion buffer (0.8x TAE buffer, 0.5% Triton X-100, 0.8 M guanidine HCl) and incubated overnight at room temperature or for 4-6 hours at 37°C. Usually after this time gels were expanded up to 2 cm and cover slip was easily peeled off. Gels were then washed once with 1x PBS, stained with 2 μM DAPI solution in PBS, washed with PBS again and expanded in deionized water. Water was changed every 30-40 minutes until the gels were fully expanded (usually 3-4 exchanges).

II.4. The analysis of the structure of unmethylated DNA in MCF10A and MCF7 cells

II.4.1. Results and discussion

II.4.1.1. The aims

Epigenetic modifications are dynamically regulated by numerous proteins in order to maintain cell metabolism. An abnormal level of various epigenetic modifications, including DNA methylation, can cause the malfunction of cells leading to diseases such as cancer. Therefore, we decided to apply the developed method of the selective labelling of unmethylated DNA in cells to compare the organisation of the unmethylated chromatin in normal and cancerous breast cells. In addition, we investigated if this method can be used to visualise the difference between DNA methylation levels in cancer and healthy cell lines.

II.4.1.2. Changes of native M.TaqI target sites methylation

In order to test if the changes of methylation level can be observed using the MTase labelling method we induced a hypomethylating agent to MCF10A and MCF7 cells. Adenosine-2',3'-dialdehyde (AdOx) inhibits the AdoHcy hydrolase enzyme, leading to

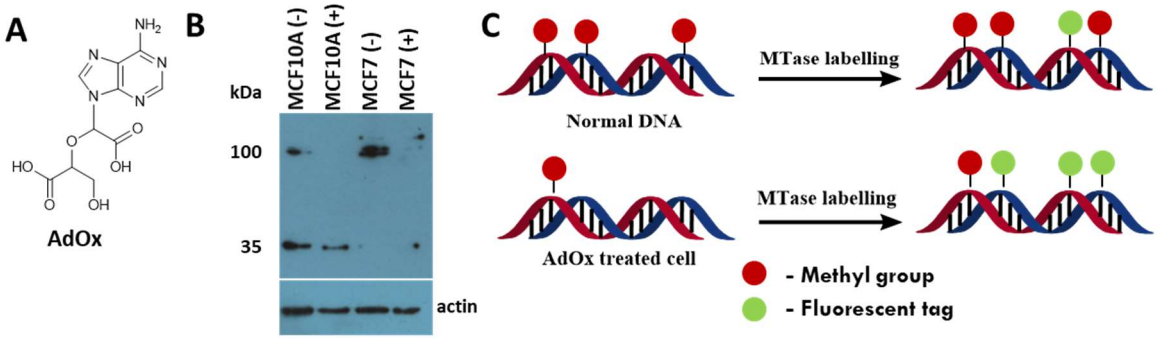


Fig. 83 Structure of AdoX molecule (**A**); Western blot of antisymmetric demethylation of arginine. Actin used as a loading standard (**B**); scheme of MTase-directed fluorescent labelling of DNA in normal and AdOx treated cells (**C**).

an accumulation of AdoHcy in the cell, which inhibits the AdoMet dependent MTases, including DNA MTases causing global hypomethylation of DNA (Fig. 83 A).[292] Thus, in the case of the TCGA sequence, if a cytosine and/or an adenine undergoes dynamic methylation/demethylation, it will preferentially adopt an unmethylated state after the inhibition of DNA MTases. Hence, upon M.TaqI-directed fluorescent labelling, more labels will be introduced to the genome by the enzyme (Fig. 83 C).

MCF10A and MCF7 cells were treated with 10 µM AdOx over 48 hours. The global level of antisymmetric dimethylation of arginine residues in proteins in cells was tested using Western Blot as a measure of the effectiveness of AdOx treatment (Fig. 83). The treatment with AdOx was successful as a weaker band at 35 kDa is present after treatment in MCF10A cells, a band at 100 kDa in both cell lines is greatly reduced upon treatment with AdOx.

Both cell lines were labelled with M.TaqI and AdoHcy-6-yne and imaged using a widefield microscopy without ExM. Both (AdOx-treated and the control) samples show clusters of labelling sites (Fig. 84 A, C, D, F 1). In the images of treated cells, many small circular foci (around 500 nm) are observed (Fig. 84 D, F 2). These were not seen in the control sample in such numbers. This is consistent with genome-wide demethylation, and a corresponding increase in M.TaqI-directed labelling. It is difficult to notice any systematic change of shape or size of the large clusters due to the low resolution of the images.

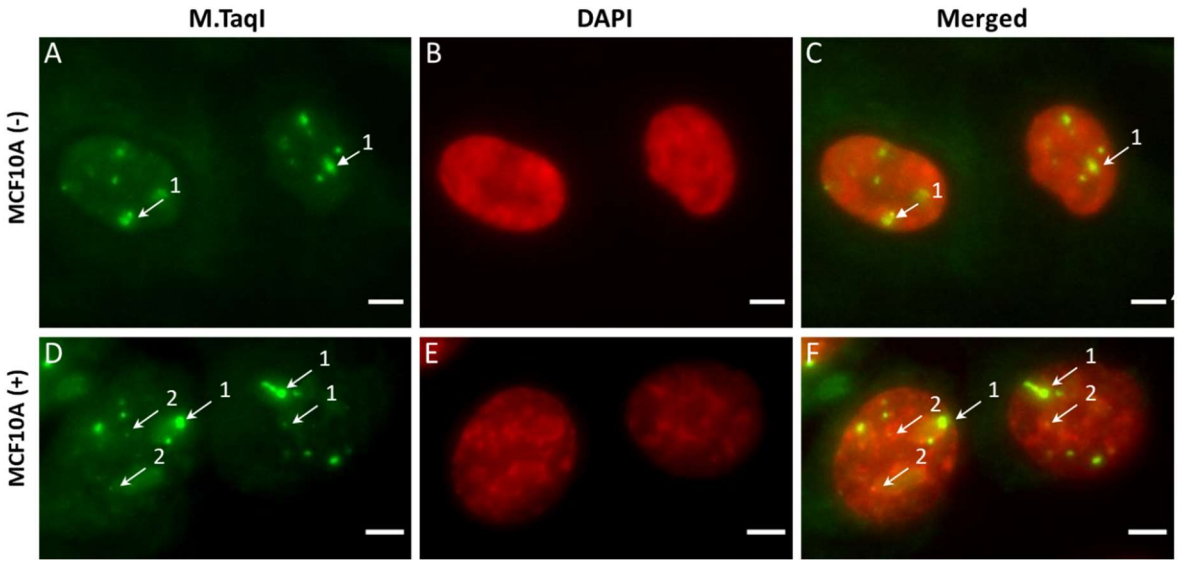


Fig. 84 Labelling of MCF10A with M.TaqI and AdoHcy-6-yne. Cells were (**A-C**) untreated (-) and (**D-F**) treated with AdOx (+). Scale bars = 5 µm, images shown as a maximum intensity Z-projection.

In both normal and cancer cells, a significant difference in the number of clusters of labels in the nuclei is observed before and after treatment with AdOx (Fig. 85). A nucleus of treated MCF10A cells contains on average 9.1 ± 2.2 foci compared to 5.8 ± 1.5 in untreated cells. The difference is smaller in case of cancer MCF7 cell line where a nucleus in the treated cell contains 9.0 ± 1.8 compared to 7.9 ± 4.6 in the untreated cell.

Although, an average nucleus of MCF7(-) contains considerably more foci than MCF10(-), there is no statistical difference between both cell lines after the treatment. This may be due to the limit of demethylation or an equilibrium between global methylation and unmethylation of DNA. However, it is quite surprising as even more

foci would be expected in MCF7 cells due to the presence of a higher chromosome copy number in these cells.

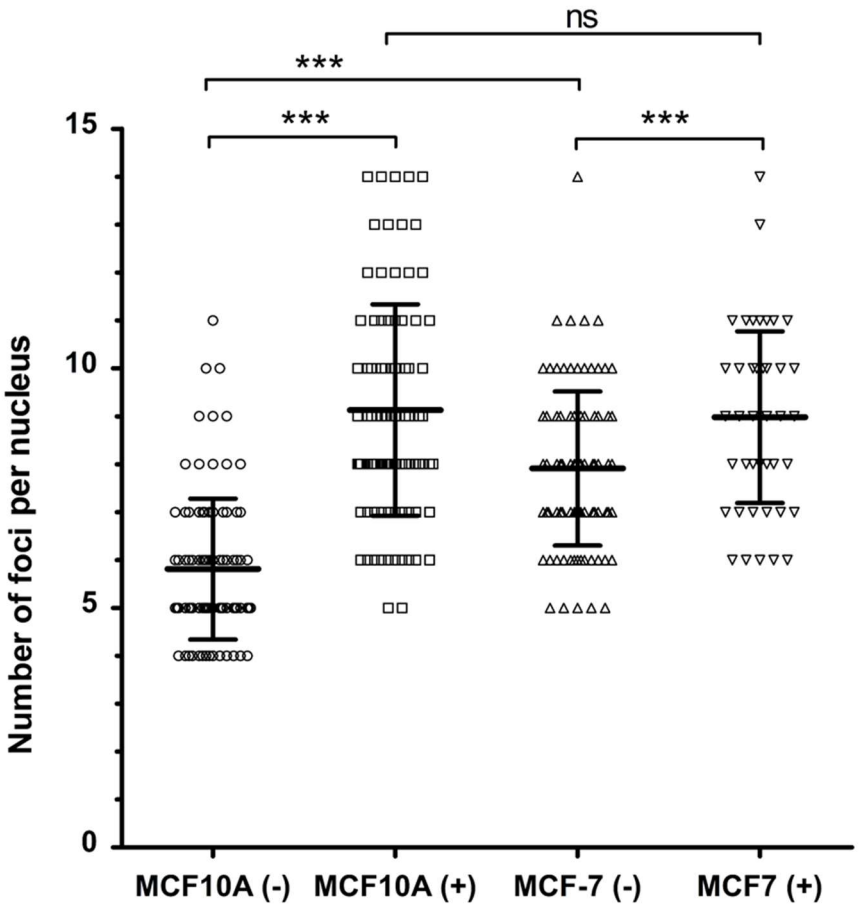


Fig. 85 Count of foci of M.TaqI target sites per nucleus in MCF10A and MCF7 cells both untreated (-) and treated with AdOx (+).p-values obtained from Mann-Whitney test: MCF10A(-) vs MCF10A(+) $p < 0.0001$; MCF10A(-) vs MCF7(-) $p < 0.0001$; MCF7(-) vs MCF7(+) $p < 0.0001$; MCF10A(+) vs MCF7(+) $p = 0.9598$. MCF10A(-) N = 100; MCF10A(+) N = 100; MCF7(-) N = 100; MCF7(+) N =53. Error bars as standard deviation

This simple analysis can be used to monitor the change of methylation of TCGA sequences but cannot be used to characterize and distinguish specific cell types. Therefore, we looked into the structure of unmethylated M.TaqI target sites regions of DNA in the nuclei and aimed to find a descriptive trait, using iSPIM images, that does not depend on the number of chromosomes present in a given nucleus.

II.4.1.3. Segmentation of M.TaqI target sites clusters

As a feature that could be used to compare M.TaqI labelled cells, we decided to use a ratio of volume occupied by M.TaqI labels in the nucleus to the volume of this nucleus. An increased amount of genetic information should increase both the volume of M.TaqI labels and the volume of the nucleus, hence the methylation level would have a greater impact on the ratio value than varying number of chromosome copies.

In order to calculate the volume of any object it needs to be identified in the image. To do this a threshold of a fluorescence intensity of background noise to the signal from the object of interest needs to be identified. This allows a segmentation of an object in a single image, which then can be analysed in three dimensions with various 3D segmentation plugins available in Fiji. Due to the relatively even staining of nuclei with DAPI and a good signal to noise ratio, the images were thresholded using the Huang method, which is based on minimizing the fuzziness of an image.[293] This method, out of 17 thresholding methods available by default in Fiji, showed the most reliable thresholding of the outer shape of the nucleus. After segmentation, the volume of nucleus, based on DAPI staining, was measured using the 'Analyze particles' function in Fiji.

We encountered a significant challenge in the reliable segmentation of the objects present in the M.TaqI channel of iSPIM images. Typically, there were three main groups of objects observed in these images. Large bright objects (>1-1.5 μm in diameter; Fig. 86 A 1), small round objects (< 1 μm in diameter; Fig. 86 A 2) and a fluorescence signal from the nucleoli (Fig. 86 A 3). Our first attempt was to apply an automatic thresholding methods available in Fiji software. All of the tested methods either overestimated or underestimated surfaces of the objects. Out of all the thresholding methods tested, after a visual assessment the most promising results

were obtained with Moments method. Some methods such as Huang or MaxEntropy are not able to distinguish any structures from the background noise. Others such as Otsu, are highly biased towards large clusters. The Moments method was able to threshold both large and small clusters. However, although it works well in preserving small objects, it struggled with more complex regions containing clusters of labels with more complex topologies. The fluorescence of the nucleoli is also recognised as an object which leads to major overestimation of both the labelled surface area and volume in three-dimensions (Fig. 86 B 3). Moreover, the small objects are not properly resolved when they colocalise with a signal from nucleoli (Fig. 86 B 2). The shape and complex structure of the large objects is also lost and overestimated and may lead to creation of massive objects as e.g. when combined with nucleoli (Fig. 86 B 1).

Hence, the heterogeneity of the labelling densities, shapes and sizes of these structures requires a complex and adaptable approach in order to derive a proper quantitative analysis of the clustering patterns. As such, we tested Ilastik, a machine-learning platform.[294] Ilastik is able to more accurately distinguish objects of interest from signal, such as that seen in the nucleoli, which we might consider to be ‘noise’. Firstly, the software is trained on a set of images where both signals of interesting objects and noise are manually entered. Then, a set of measurements that characterize the system best is chosen. The trained programme is then used to extract the information on features of interest from the images. The machine-learning based approach allows us to exclude most of the nucleoli and extract small objects when colocalised with nucleoli (Fig. 86 C 2) A tremendous improvement was made in the large clusters regions. A very detailed shape of these objects is preserved after thresholding which allows more precise surface measurements but also preserves the

information about shape and can resolve some of the smaller objects in a close proximity (Fig. 86 C 1).

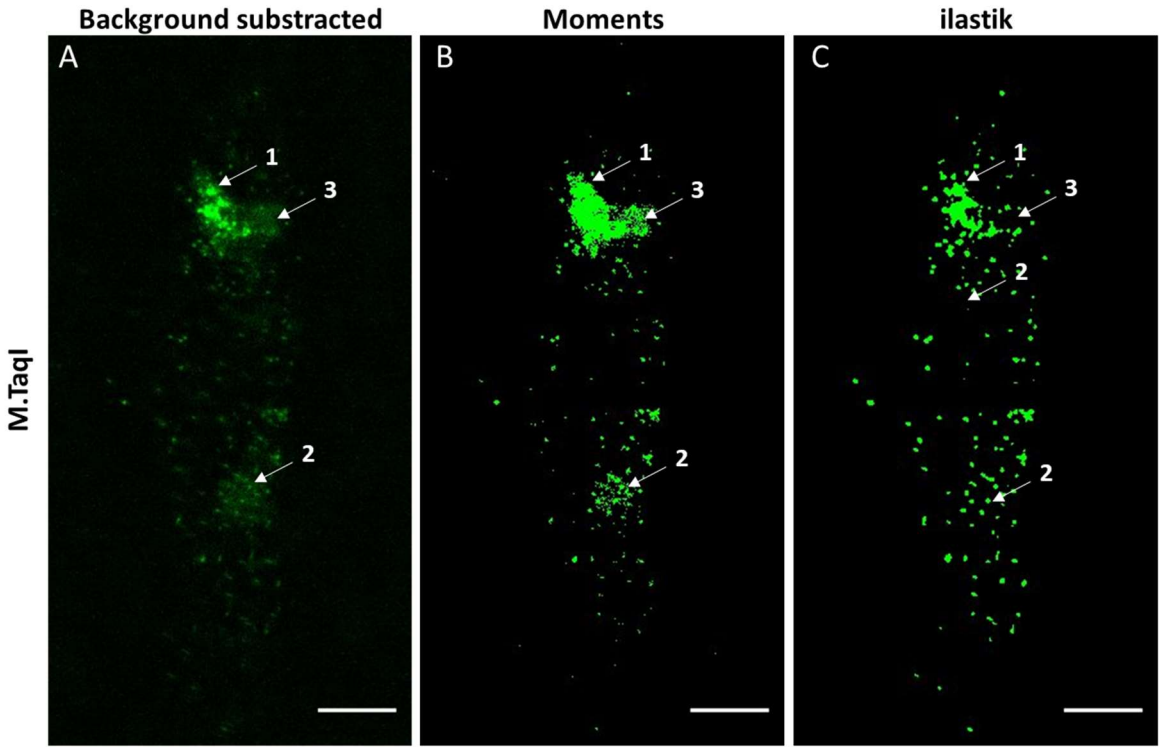


Fig. 86 Comparison of thresholding methods using inbuilt Moments in Fiji and machine-learning software ilastik. Single plane of a M.TaqI labelled cell after background substraction (**A**), thresholded using Moments method (**B**) and thresholded using ilastik software (**C**). Moments method overestimates large cluster reagions as well as nucleolar regions by thresholding background noise from non-specifically attached labels. Ilastik is able to threshold signal from both the large clusters and small foci and reduces the background signal.

We determined the volume of segmented objects using the 3D Manager plugin in Fiji and the volume of each object was measured. For each nucleus ratios of the total volume of fluorescent objects in the M.TaqI channel to the total volume of the nucleus was calculated and data from a set of nuclei, for each cell line, was analysed (Fig. 87). On average, the total volume occupied by labelled, unmethylated regions of M.TaqI target sites in the genome is increased in the cancer-derived MCF7(-) (6.1 % ± 1.8) compared to healthy MCF10(-) (4.4 % ± 1.8) cells. As expected, M.TaqI labelled DNA occupies a larger volume of the nucleus in AdOx treated MCF10A cells (6.3 % ± 1.8)

compared to the untreated counterparts (4.4 % ± 1.8). There is no significant difference between the total volume of labelled regions of M.TaqI target sites in MCF7(-) and

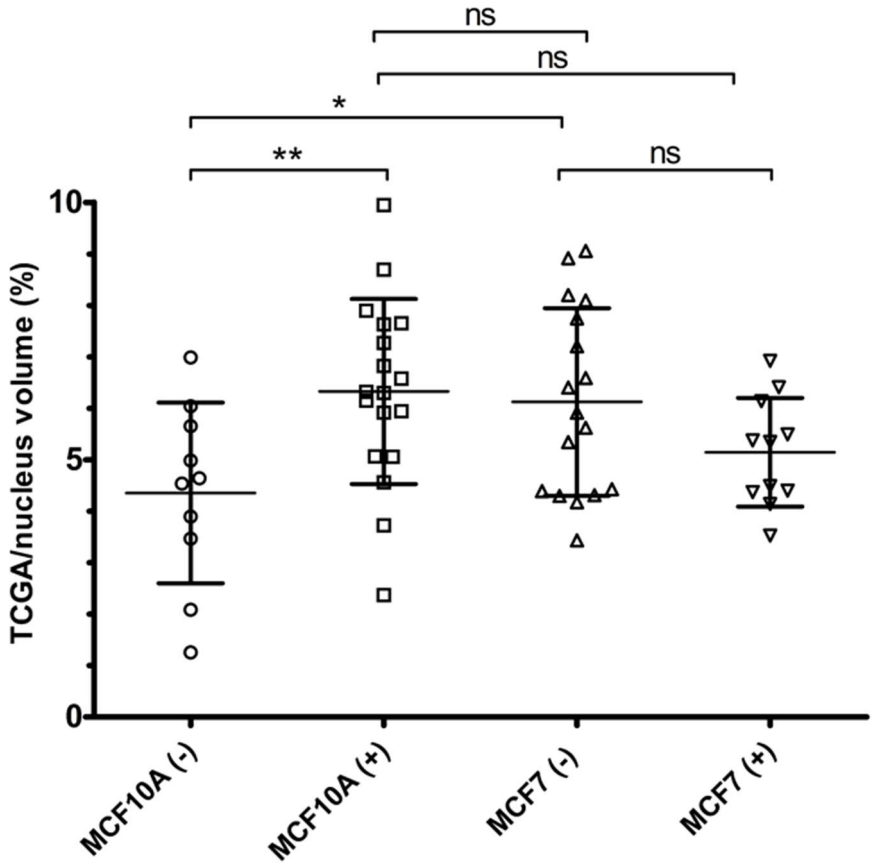


Fig. 87 Volumetric ratio of labels of M.TaqI target sites per nucleus in MCF10A and MCF7 cells both untreated (-) and treated with AdOx (+).p-values obtained from unpaired t-test: MCF10A(-) vs MCF10A (+) p = 0.0094; MCF10A(-) vs MCF7(-) p = 0.0207; MCF7(-) vs MCF7(+) p = 0.1200; MCF10A (+) vs MCF7(-) p = 0.7434; MCF10A (+) vs MCF7(+) p = 0.0591. MCF10A(-) N = 10; MCF10A (+) N = 18; MCF7(-) N = 17; MCF7(+) N =11. Error bars as standard deviation

MCF7(+) cells. In the applied volumetric method the impact of a number of chromosome copies in the nucleus is decreased compared to the spot counting method previously applied. Hence, a volume fraction of M.TaqI labels is not sufficient to distinguish between the two MCF7 samples (AdOx-treated and control). This is likely due to the fact that there is already an increased level of unmethylation of regions of M.TaqI target sites in MCF7(-), compared to MCF10A(-), since cancer leads to global hypomethylation of the genome. In addition, there is not a significant difference

between MCF10A(+) and both MCF7(-) and MCF7(+) cells which would suggest that a maximum level of unmethylation of regions of M.TaqI target sites was reached in MCF10A(+) sample and is similar to both hypomethylated cancer-derived MCF7 samples, which was at least distinguishable with this method. This can be an effect of a small samples size and simplicity of this approach it might not be sensitive enough to distinguish small differences between these cell lines (if there is any). Nevertheless, this approach is promising in comparative analysis of methylation level in various cell types.

In order to study the structure of unmethylated regions of M.TaqI target sites in MCF10A and MCF7 cells, we decided to use a statistical analysis of other characteristics of the objects obtained during volumetric measurements.

II.4.1.4. Statistical analysis of three-dimensional architecture of unmethylated regions of M.TaqI target sites

Unmethylated regions of M.TaqI target sites of the genome form structures with a broad range of sizes and shapes. The size of identified objects spans across 4 orders of magnitude with the largest with a volume of hundreds of μm^3 in expanded samples (Fig. 88). The distribution of volumes of M.TaqI labelled objects for all analysed cell lines was plotted using a Kernel smooth density estimation in Origin 2020. Such estimation smooths a discrete distribution by taking an average value of neighbouring data. In the result, an estimated continuous distribution is presented. For all samples the distribution is similar with a vast majority of objects with a volume between 0.1-1 μm^3 (Fig. 88 left panel). Although this peak is shifted depending on the cell line, the consistency of thresholding of such small objects in each cell may differ. Therefore, we

decided to focus on larger structures where size and feature count is more accurately determined. For volumes above 10 μm^3, the distributions of feature volumes have characteristic profiles (Fig. 88 right panel). It is possible that these large features are intimately linked to the genome biology in the sample, and hence, we explored the analysis of this subset of features with volumes larger than10 μm^3, specifically.

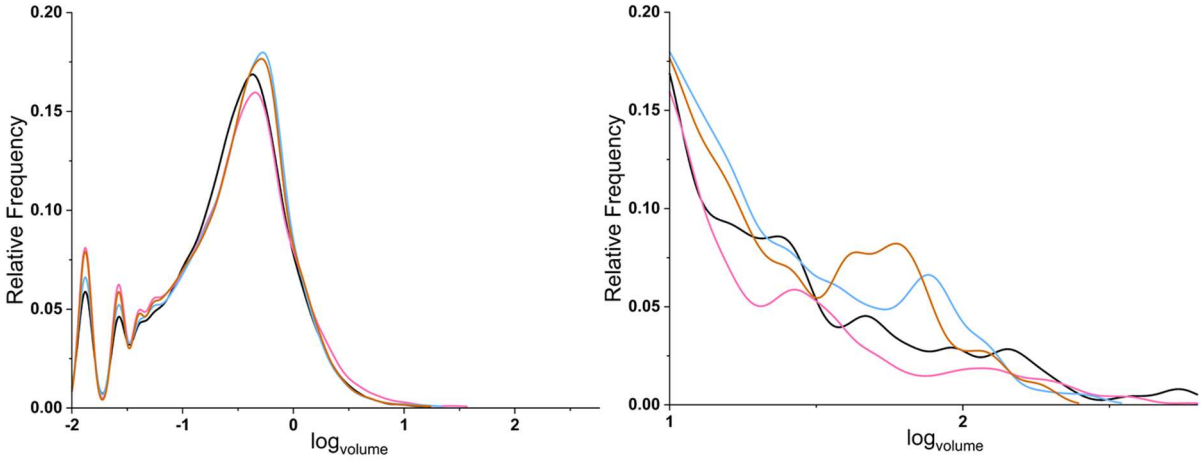


Fig. 88 Distribution of volumes M.TaqI labelled objects (left) and a distribution of volumes above 10 μm^3 (right) in MCF10 (-) – black, MCF10A (+) – light pink, MCF7(-) – light brown, MCF7(+) – light blue. Logarithm of volume in μm^3.

A distribution of sphericity feature, which is calculated as the cube root of the normalized ratio between the surface and the volume, of these objects was calculated (Fig. 89 A-D).

The objects in MCF10A(-) cells form two groups: S - objects smaller than ~40 μm^3, with a maximum of a peak at ~16.5 μm^3; L - larger than ~40 μm^3, with a maximum at ~132 μm^3 (Fig. 89 A). After the treatment with AdOx, two groups (S'* and S''*) can be identified in the lower volume, with maxima at ~12.5 μm^3 and ~28 μm^3 respectively, and a L* group of largest objects with a maximum at ~112 μm^3. In addition, more objects in AdOx treated MCF10A cells tend to be less spherical compared to ones in untreated cells (Fig. 89 B). Two examples of large objects with high sphericity value in

MCF10A(-) and smaller objects with a lower value of sphericity in MCF10A(+) are presented in Fig. 89 E and F respectively.

In MCF7 cells, both before and after treatment, two groups of objects can be identified. S and S* groups of smaller objects in both samples have maxima at a similar volume value, ~13.5 μm^3 in MCF7(-) and ~14.5 μm^3 in MCF (+). The maximum of the L* group is significantly shifted towards greater volumes in AdOx treated MCF7 cells (~72 μm^3 compared to ~51 μm^3 in MCF7(-)). In addition, the sphericity of M.TaqI labelled objects is dramatically decreased in AdOx treated MCF7 cells.

A decreased level of the global methylation causes a disruption to the integrity of the unmethylated M.TaqI target sites structures in both MCF10A and MCF7 cells. In particular, in MCF7 cells, a large decrease of sphericity as well as a significant increase of the volume of large objects suggest that the structure of those was affected the most. However, in MCF10A cells, the treatment affected the S group the most rather than the L group. The origin of the two distinct groups – S'* and S''* - is unknown. The split-like pattern of the S group into two with a lower and a higher volume and no significant change in the volume distribution of the L group, could be a result of two opposite mechanisms that either increase or decrease the sphericity significantly of these objects thus, decreasing and increasing the volume. However, as the S group is larger than L group, the split pattern should be reflected in a distribution of sphericity. A lack of such pattern would suggest that after the treatment, the S group shifted, presumably towards larger volumes with a maximum at 28 μm^3, and the second group, with a maximum at 12.5 μm^3 is simply formed from objects smaller than 10 μm^3 before the treatment.

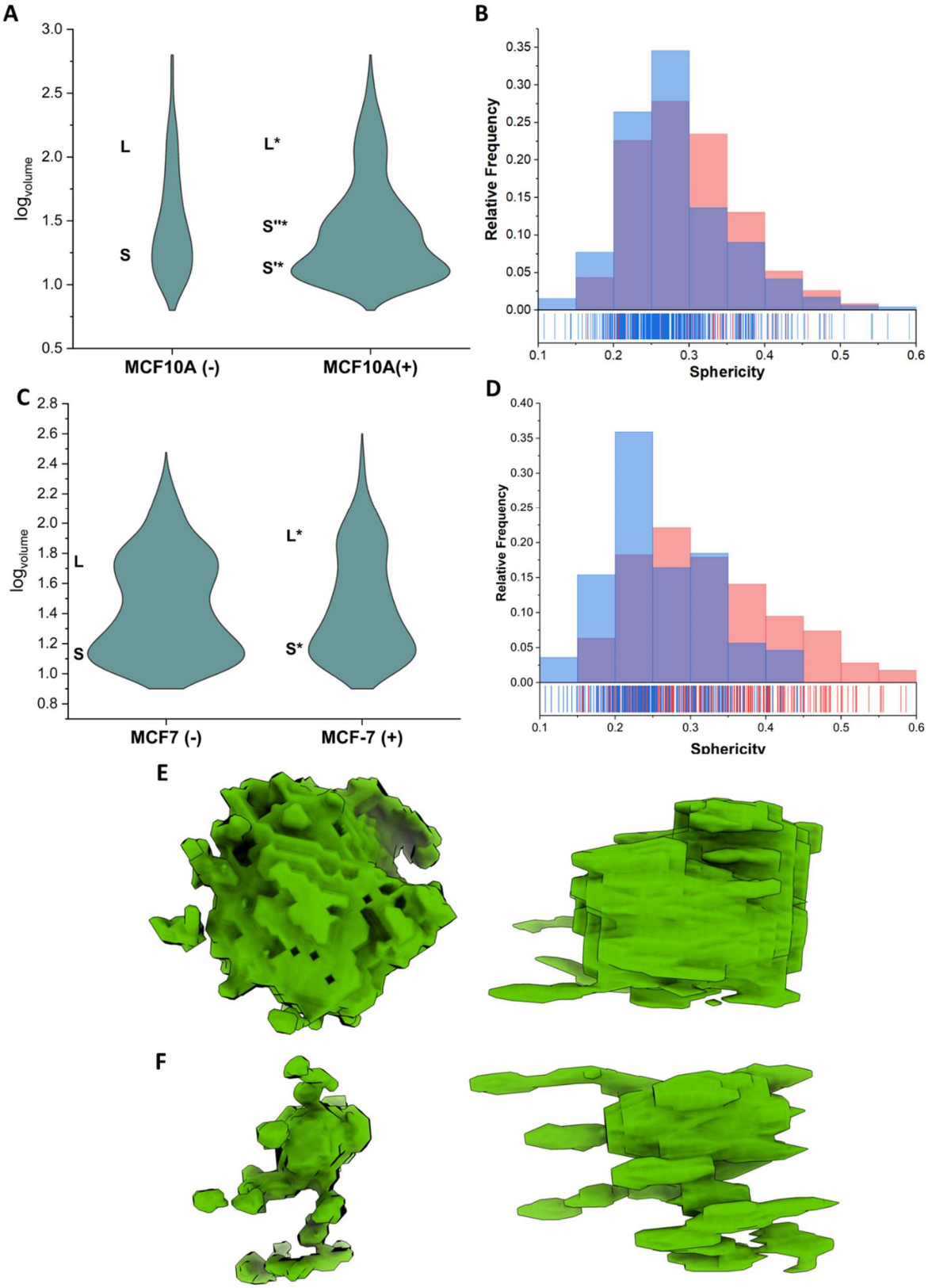


Fig. 89 Comparison of volume and sphericity distribution of M.TaqI labelled objects between: MCF10A (-) versus MCF10A (+) cells and MCF7 (-) versus MCF7 (+). Violin plot of volumes of M.TaqI labelled objects in MCF10A(-) and MCF10A (+) – **A**, in MCF7(-) and MCF7(+) – **C**. Histogram with a rug of sphericity of unmethylated objects of M.TaqI target sites in MCF10A(-) (red) and MCF10A (+) (blue) – **B**, MCF7(-) (red) and MCF7(+) (blue) – **D**; a top view (left) and a side view (right) of an object in expanded MCF10A(-) cell (volume 146 μm^3 and 0.42 sphericity) – **E**, an object in MCF10A (+) cell (volume 32 μm^3, sphericity 0.28) – **F**. Logarithm of volume in μm^3.

The analysis of both distributions of volume and sphericity is able to distinguish between tested cell lines both before and after treatment with AdOx. Objects of the L group in MCF7 are much smaller compared to MCF10A (~2.6 times before treatment and ~1.5 times after treatment) Fig. 90 A, C. Interestingly, objects in MCF7 tend to be more spherical compared to MCF10A, whereas the M.TaqI labelled regions after AdOx treatment are less spherical in MCF7 (Fig. 90 B, D).

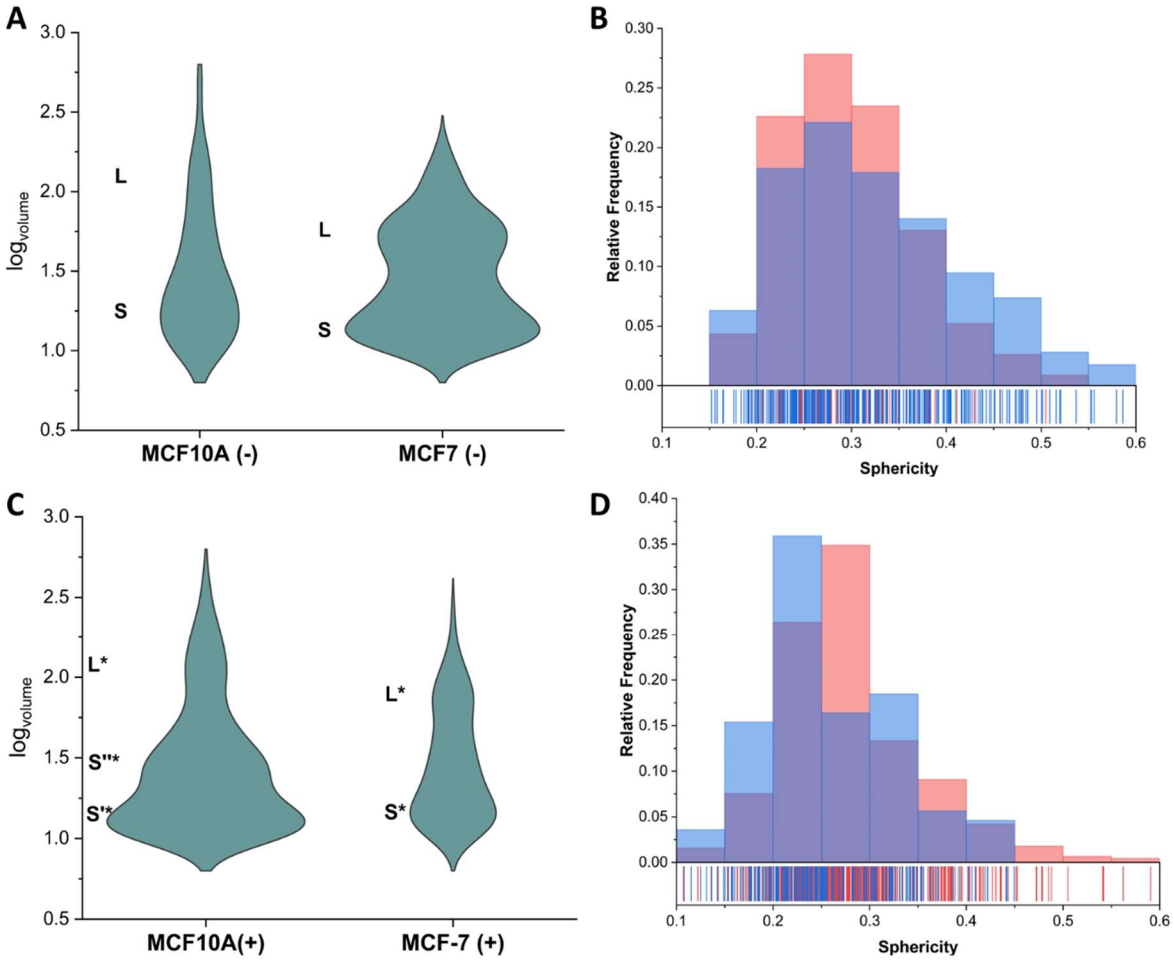


Fig. 90 Comparison of volume and sphericity distribution of M.TaqI labelled objects between: MCF10A (-) versus MCF7 (-) cells and MCF10A (+) versus MCF7 (+). Violin plot of volumes of M.TaqI labelled objects in MCF10A(-) and MCF7(-) – **A**, in MCF10A (+) and MCF7(+) – **C**. Histogram with a rug of sphericity of unmethylated objects of M.TaqI target in MCF10A(-) (red) and MCF7(-) (blue) – **B**, MCF10A (+) (red) and MCF7(+) (blue) – **D**. Logarithm of volume in μm^3.

The results of this statistical analysis can be visualised at the single cell level where the objects in AdOx treated samples (Fig. 91 B, D) are generally, less spherical than in untreated counterparts (Fig. 91 A, C).

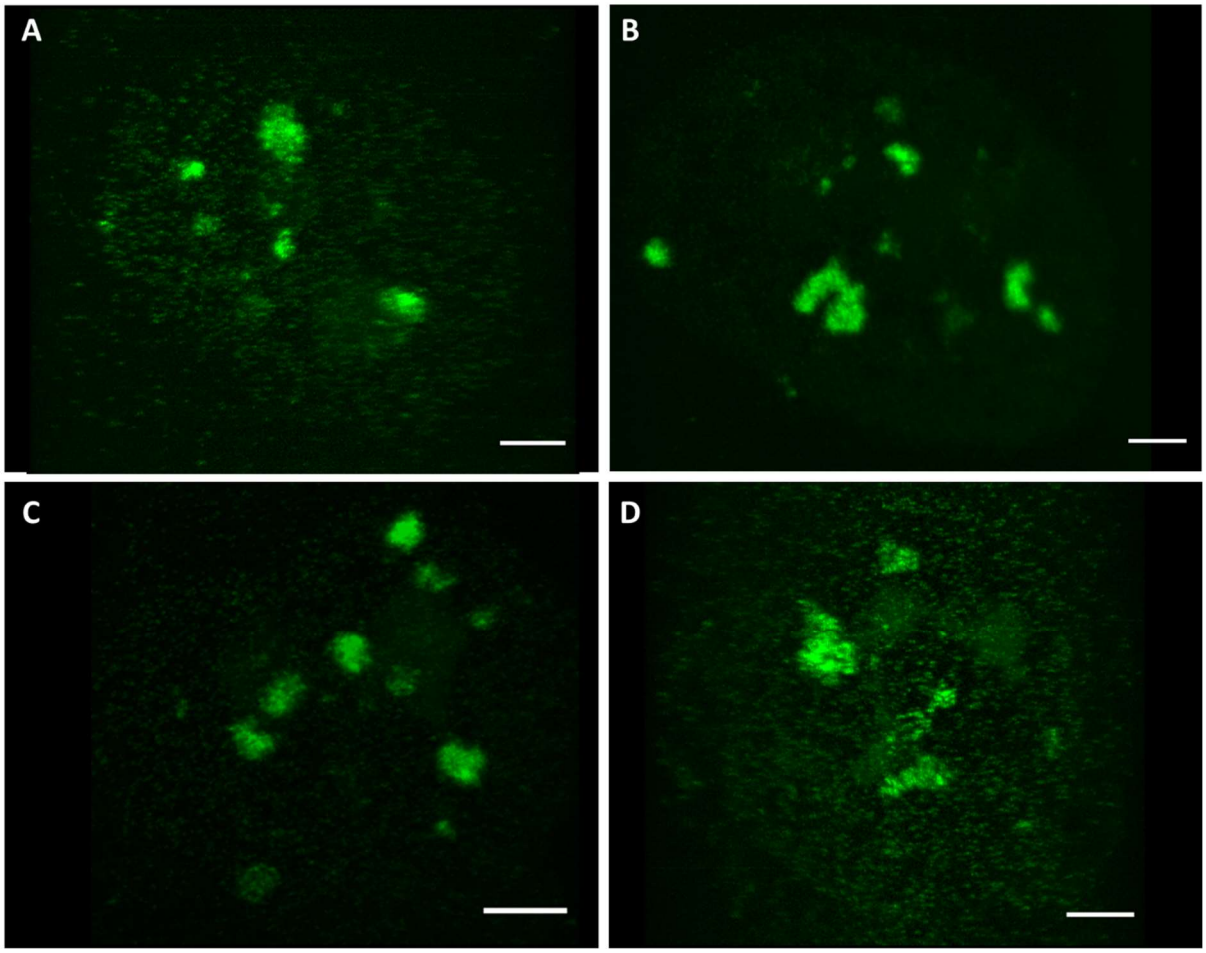


Fig. 91 Three-dimensional projection of a MCF10A(-) – **A**, MCF10A (+) – **B**, MCF7(-) – **C**, MCF7(+) – **D** expanded cells. Scale bars = 10 μm.

Such a loss of a compact structure of unmethylated M.TaqI target sites clusters, especially in cancer MCF7 cells, upon induced hypomethylation is similar to the studies of structure of transcription factories in *HeLa* cells in nucleolar regions (see Fig. 8).[115] In addition, as dense regions of M.TaqI target sites were identified in accessible repetitive regions of chromatin which are rich in transcription factor recognition sequences, it is probable that the observed clusters play a functional role in these transcription factories. In the future studies, a possible distribution changes of M.TaqI labelled objects and transcriptionally important proteins, such as RNA polymerase,

could give an insight into the importance and role of the chromatin regions rich in M.TaqI target sites in transcription.

In cancer cells, unlike in normal counterparts, the global methylation level is not significantly altered but rather the distribution of unmethylated sites across the nucleus.

II.4.2. Conclusion

We applied the developed method of methyltransferase-directed fluorescent labelling of unmethylated DNA in cells to study the changes in the methylation level of 5'-TCGA-3' containing regions of DNA. We induced global hypomethylation in MCF10A and MCF7 cells by treatment with non-specific methylation inhibitor – AdOx. We showed that a simple analysis of number of foci in the nucleus can be applied to measure the relative change of methylation level of M.TaqI target sites upon treatment with AdOx in both healthy MCF10A and cancer MCF7 cells. However, this method is not appropriate for comparison of different cell lines due to varying number of copies of DNA in each.

To tackle this we used a volumetric ratio of M.TaqI labelled structures to the nucleus as a descriptor. We applied a machine-learning approach to accurately segment the M.TaqI labelled structures. We found that a simple volumetric ratio can be applied to track relative changes in methylation level of M.TaqI target sites in healthy MCF10A cells upon treatment as well as to compare different cell line prior treatment (MCF10A and MCF7). However, it is not possible to discriminate between treated and untreated cancer cell line (MCF7) using this approach.

Since the images show clear differences between the cluster shapes in MCF7/MCF10A cell lines, we performed a comparative analysis of volume and

sphericity distributions of objects larger than 10 μm3 in expanded samples. We identified two groups of M.TaqI labelled objects in both cell lines with distinct average sizes. We found that in MCF10A cells, upon treatment with AdOx, a third group of average size M.TaqI labelled objects can be identified with small changes of average sized in two other groups. On the other hand, in cancer MCF7 cells, upon treatment with AdOx, only two groups can be identified with the average size of objects in each group increased. In both cell lines sphericity of analysed objects decreases in AdOx-treated cells. This approach not only enabled analysis of changes in organisation of M.TaqI target sites as a result of changes in methylation level in the same cell line but also is able to show the difference between the cell line. In fact, MCF7 objects in both identified groups are on average smaller than in healthy cells and are more compact.

Although, the origin of such changes is not clear we suggest that these regions can play a role in transcription factories, as a high density of M.TaqI target sites was predicted in regions rich in transcription factor recognition sequences.

II.4.3. Materials and methods

AdoX treatment

The AdoX was added to the medium at 10 μM concentration and cells with AdoX containing medium for 24 hours. Then cells were seeded in 24-well plate on poly-L-lysine coated glass cover slips, at 40000 cells per well concentration and incubated for another 24 hours with AdoX containing medium.

Western blot

MCF10A and MCF7 (both AdOx treated and controls) were grown and treated as previously described with seeding densities (MCF10A – 3.33 x 10^5 cells in 10 cm^2

culture dish; MCF7 – 6.66 x 10^5 cells in 10 cm^2 culture dish). Cells were harvested, and washed with 1xPBS. Pellets were resuspended in 150 µL of NP40 lysis buffer with inhibitors (Tris-HCl pH 8.0, 150 mM NaCl, 50mM, 1 mM EDTA, 1% NP40, 1 mM Na_3VO_4, 50 mM NaF, 1 mM b-glycero-phosphate,1 mM phenylmethylsulfonyl fluoride (PMSF), 10 µg/mL leupeptin, 10 µg/mL aprotinin). The samples were sonicated with a probe sonicator for 5 seconds at 25% power. The samples were then incubated on ice for 15 minutes, spinned at maximum speed for 10 minutes at 4°C. Supernatant was transferred to a new 1.5 mL tube and protein concentration quantified with Bradford Assay according to manufacturer's instructions (BioRad). 50 µg of Protein lysates were resolved on 10% SDS-PAGE gel (125 V for 1.5 hour), transferred on to PVDF, blocked with 5% milk in TBTS for 1 hour and incubated with primary antibody overnight at 4°C (Asymmetric Di-Methyl Arginine Motif [adme-R] MultiMab™ Rabbit mAb mix from CST no. 13522S; 1:1000), washed three times with 1xTBTS for 10 minutes each followed by incubation with HRP-linked secondary antibody (Goat Anti-Rabbit IgG Antibody, HRP-conjugate; 1:5000) for 1hr at room temperature. The signal was detected using ECL western blotting substrate (Pierce).

III. Conclusions and future perspectives

We developed an efficient, versatile and protecting-group-free synthesis of N^6-substituted AdoHcy/AdoMet analogues. We applied this method to create a library of AdoHcy analogues bearing a variety of functional groups including azide, alkyne, carboxyl and amine in the 6 position of adenine. Synthetic AdoHcy analogues show potential inhibitory properties against AdoHcy/AdoMet dependent enzymes including methyltransferase enzymes which are considered as therapeutic targets for cancer. In addition we show that such AdoHcy analogues can be used in the synthesis of functionalised AdoMet analgoues. We hope that a method for introduction of plethora of modification in the AdoMet structure, which can be further functionalised will find application in studying of epigenetic regulation with AdoMet dependent enzymes.

We applied two methyltransferase enzymes (M.TaqI and M.MpeI) and synthetic AdoMet analogues to site-selectively label accessible and unmethylated DNA in cells. We then used widefield and light sheet microscopy to study the organisation of unmethylated DNA in healthy and cancer models of breast epithelial cells.

We found that M.TaqI target sites in both cell lines are organised in clusters found across the nucleus. Theoretical analysis suggests that these clusters are likely rich in transcription binding sites and hence, they could be a part of transcription factories. In addition, regions of the genome with high density of M.TaqI target sites are rich in repetitive sequenced including microsatellites, which are often used as a footprint in diagnosis of cancer. Light sheet microscopy allowed us to visualise the three-dimensional structure of such clusters in individual cell. We showed that a simple statistical analysis of number of foci in cells can be applied to detect the change of

methylation level of labelled sites. A more in-depth analysis of structure of such clusters using light sheet microscopy showed that M.TaqI labelled objects are on average smaller and more compact in cancer cells. The treatment of both cell lines with methylation inhibitor (AdOx) induced the loss of sphericity of these clusters in both cell lines as well as changes in the size was observed.

We aim to study M.TaqI labelled regions in cells in more details including their correlation with proteins important in methylation of DNA or proteins, in transcription and chromatin organisation e.g. CTCF, RNA polymerase.

Moreover, we show that a distribution of volumes and sphericity of the M.TaqI labelled objects can be used to characterise and distinguish both of studied cell lines. In the future studies, we hope to study a variety of different cell lines in order to identify features that can be used to characterise each population and possibly use it in detection of various conditions including cancer.

We applied the labelling and imaging method to obtain, what we believe is a very first three-dimensional image of the architecture of accessible unmethylated CpG regions of the genome. For this we used a bacterial methyltransferase, M.MpeI. We show that, the unmethylated CpG sites are organised in a complex globular network-like structure in cells. We also were able to identify labelled foci in nucleoli which we think consist of ribosomal DNA. In addition, labelled extranuclear objects were identified in both cells line which are likely mitochondrial DNA. In cancer cells (MCF7) we also identified an increased amount heavily unmethylated extranuclear DNA which can consist of extrachromosomal circular DNA rich in oncogenes and important in tumour progression.

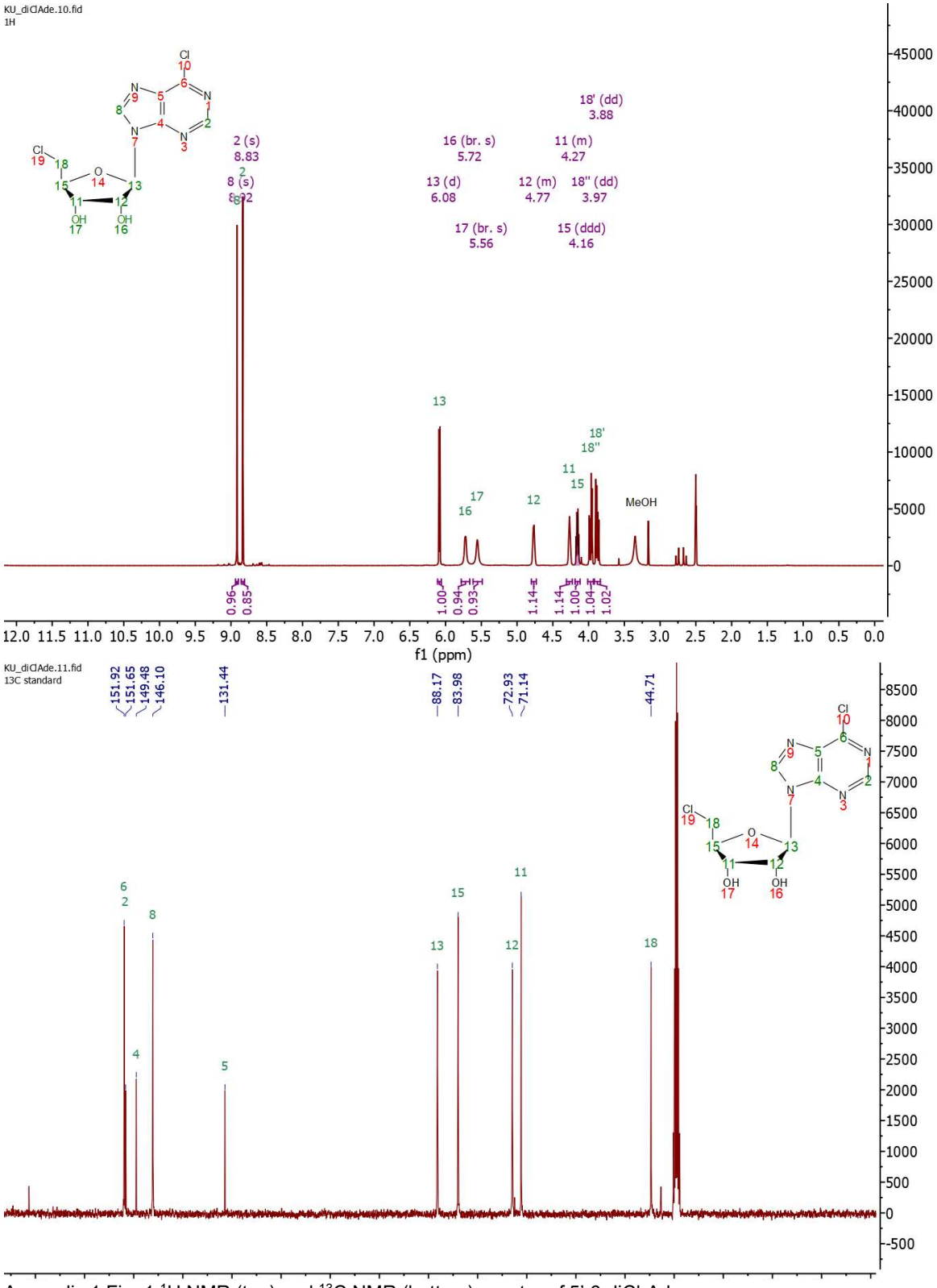
In future studies, we would like to focus on studying the changes of methylation level and organisation of three-dimensional structure of CpG sites depending on various condition e.g. different points of cell cycle, induced hypo- and hypermethylation, modulation of expression of various enzymes important in chromatin organisation and methylation.

In order, to further validate the method and identify visualised structures we are going to sequence DNA labelled according to our method. We believe this will show how selective is the labelling with both methyltransferases and help to link observed objects with specific regions in the genome.

We believe that due to importance of DNA methylation and chromatin organisation, this method will find a wide application in epigenetic studies where both of these aspects are important e.g. in early development, ageing or cancer.

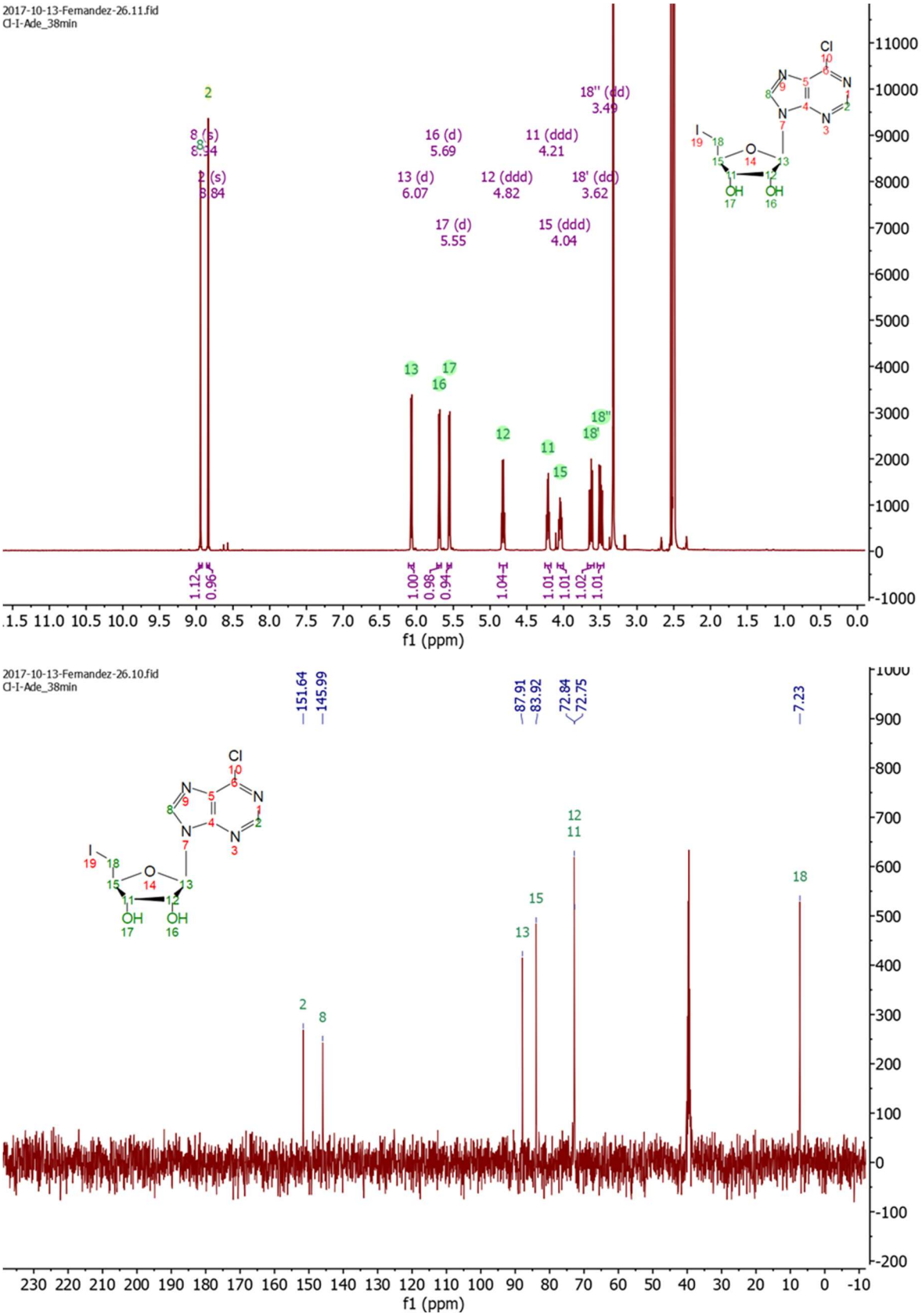
IV. Appendix A – Compound characterisation

5',6-diCl-Ade



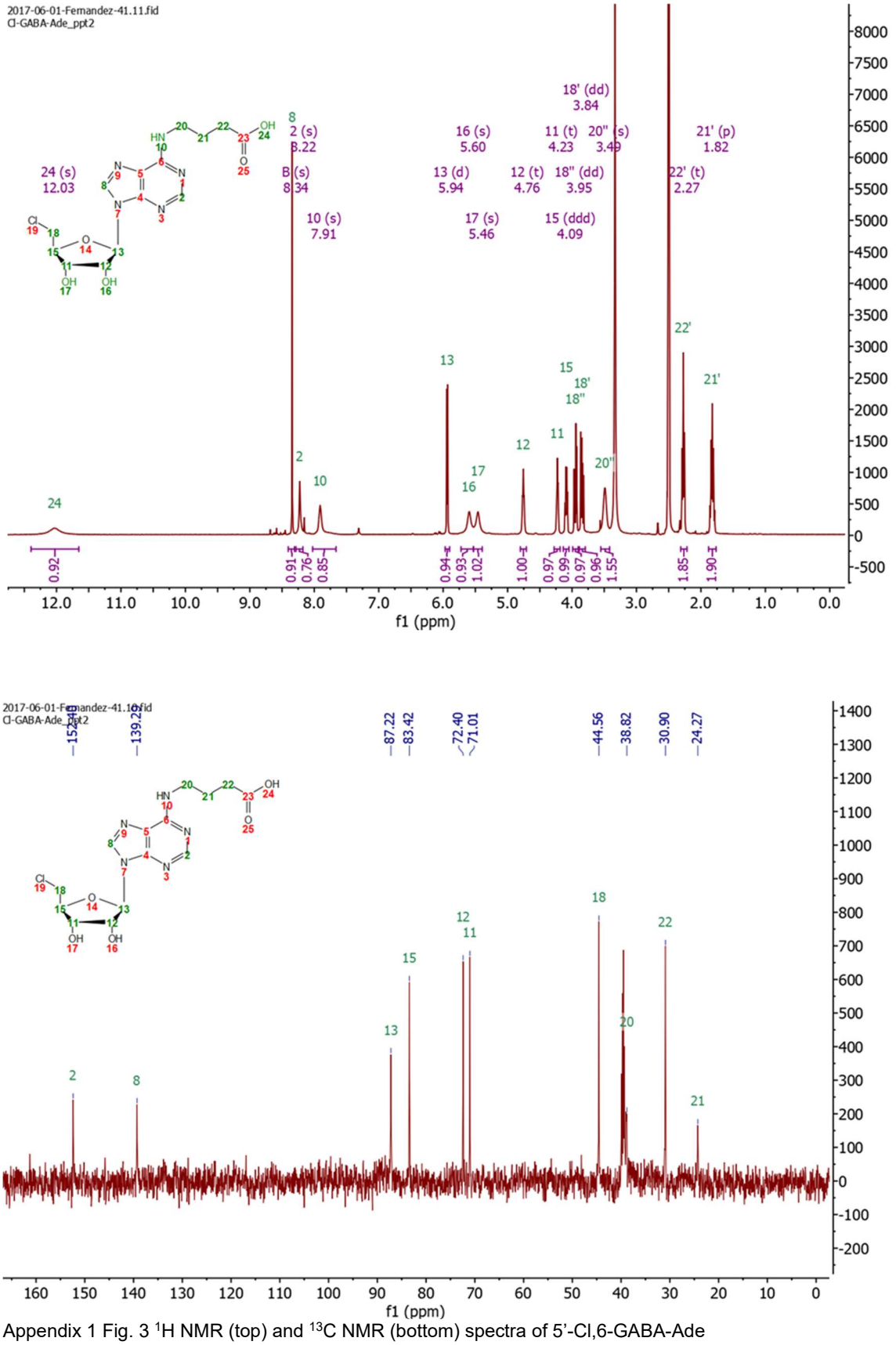
Appendix 1 Fig. 1 ^{1}H NMR (top) and ^{13}C NMR (bottom) spectra of 5',6-diCl-Ade

5’-I,6-Cl-Ade



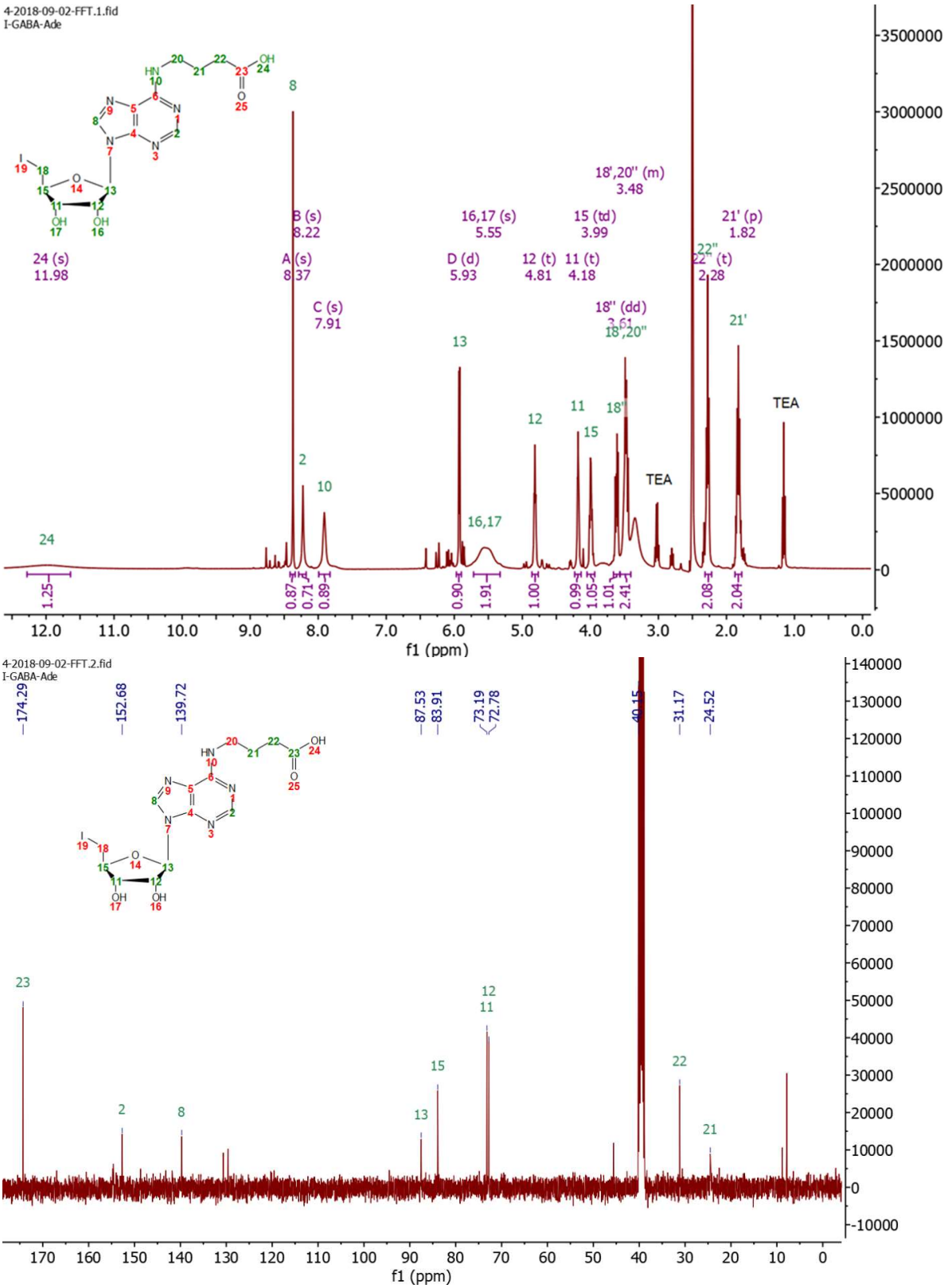
Appendix 1 Fig. 2 ^{1}H NMR (top) and ^{13}C NMR (bottom) spectra of 5’-I,6-Cl-Ade

(5’-Cl,6-GABA-Ade)



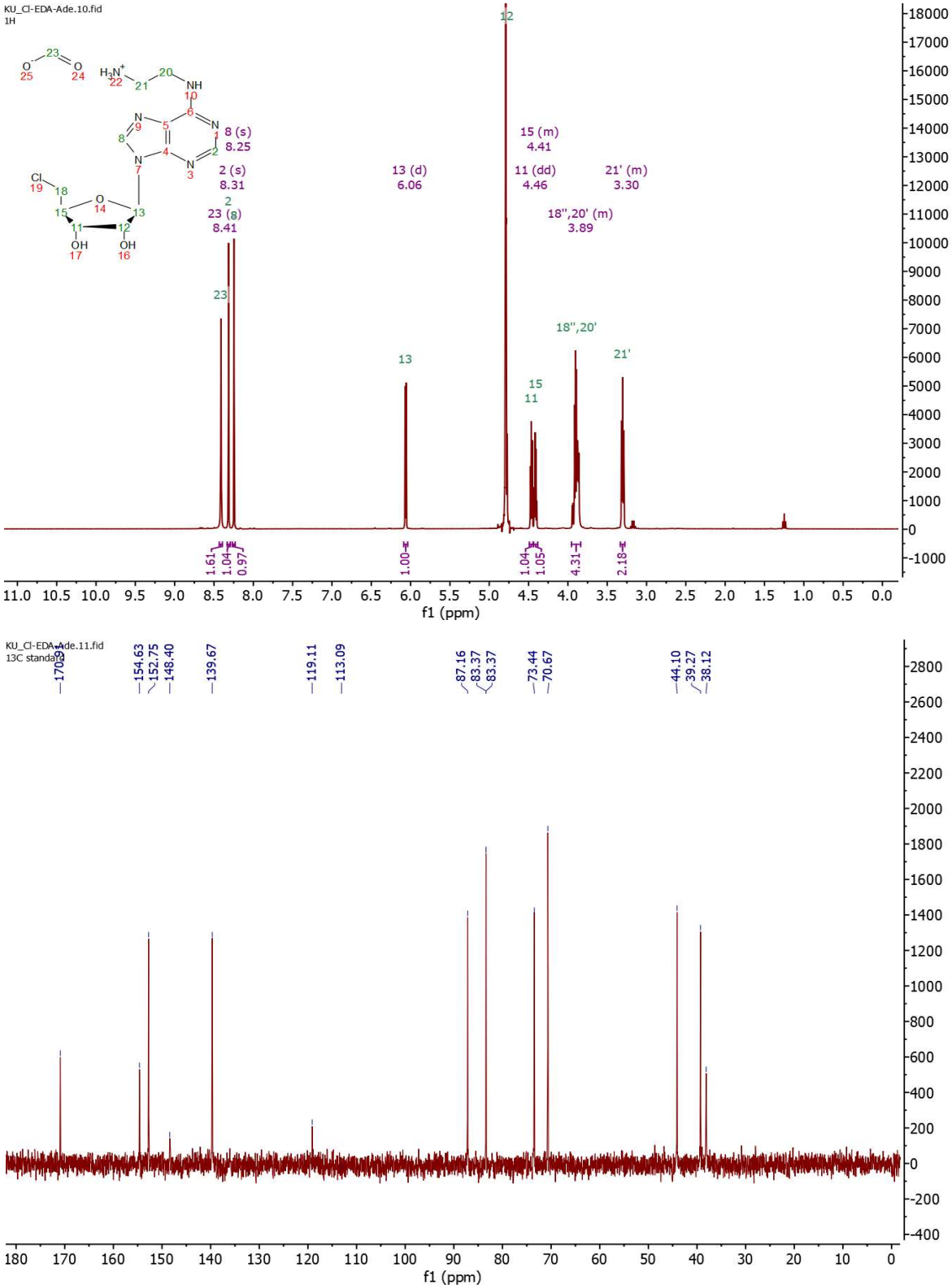
Appendix 1 Fig. 3 ^{1}H NMR (top) and ^{13}C NMR (bottom) spectra of 5’-Cl,6-GABA-Ade

5’-I,6-GABA-Ade



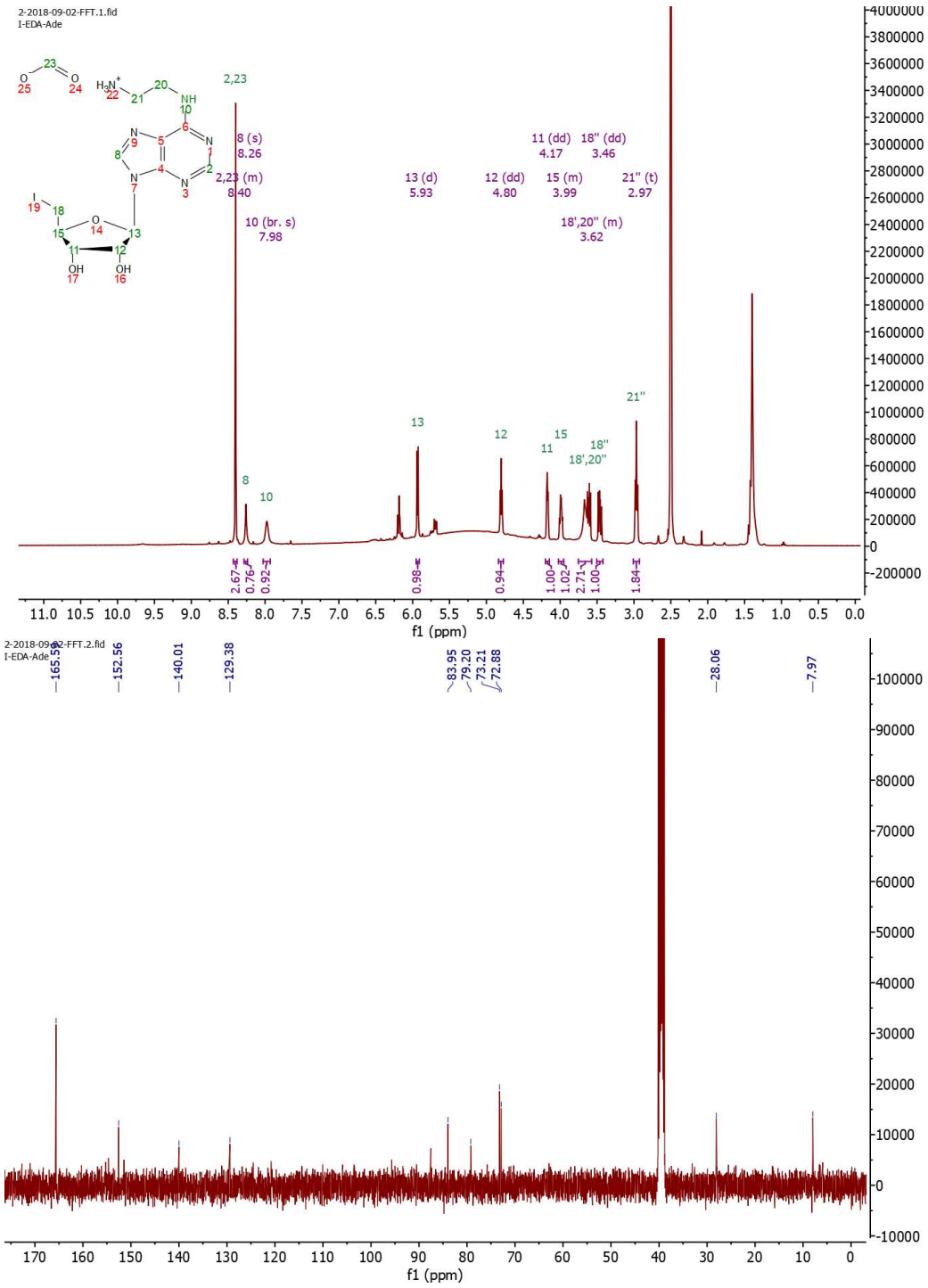
Appendix 1 Fig. 4 ^{1}H NMR (top) and ^{13}C NMR (bottom) spectra of 5’-I,6-GABA-Ade

5’-Cl,6-EDA-Ade



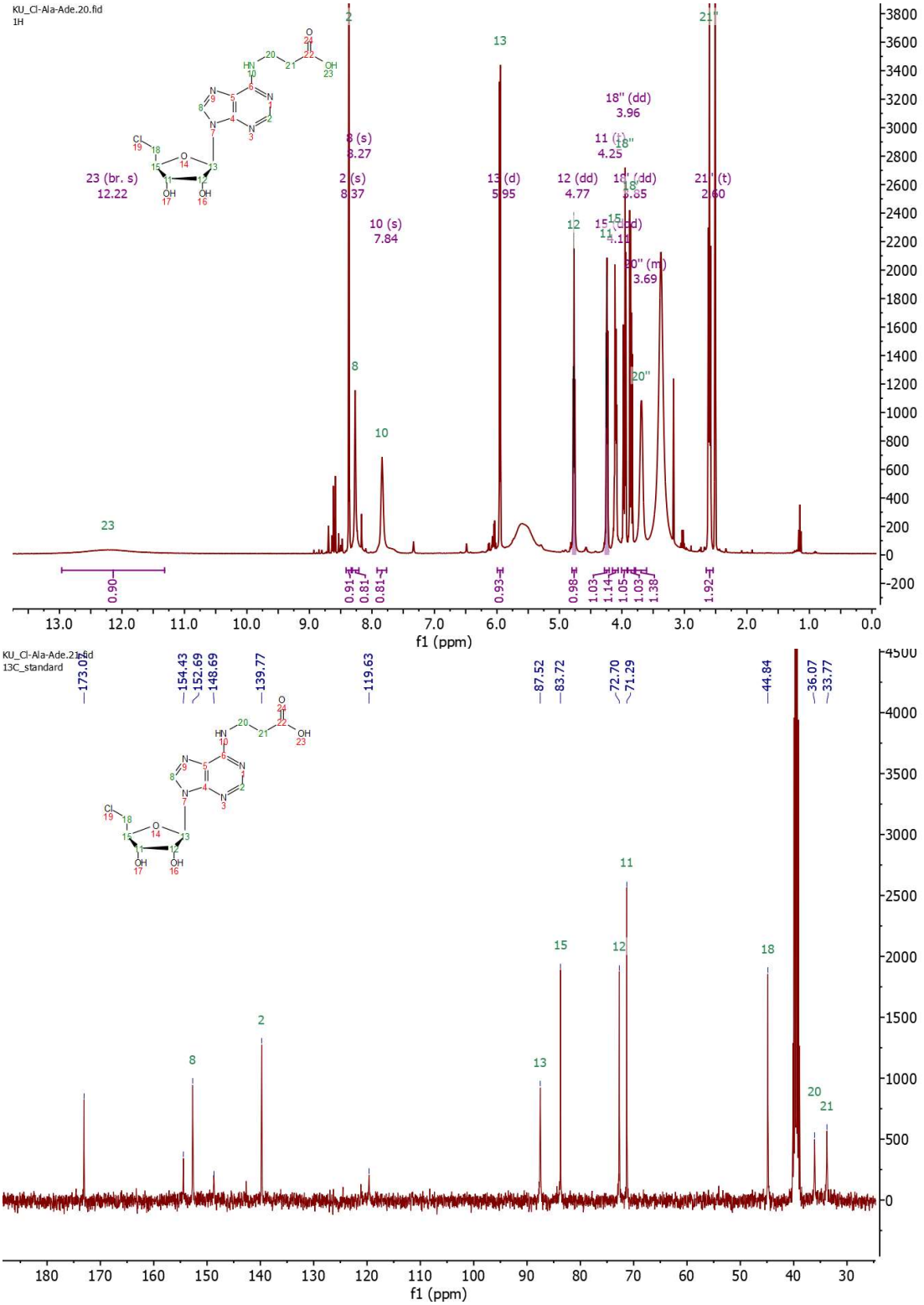
Appendix 1 Fig. 5 ^{1}H NMR (top) and ^{13}C NMR (bottom) spectra of 5’-Cl,6-EDA-Ade

5’-I,6-EDA-Ade



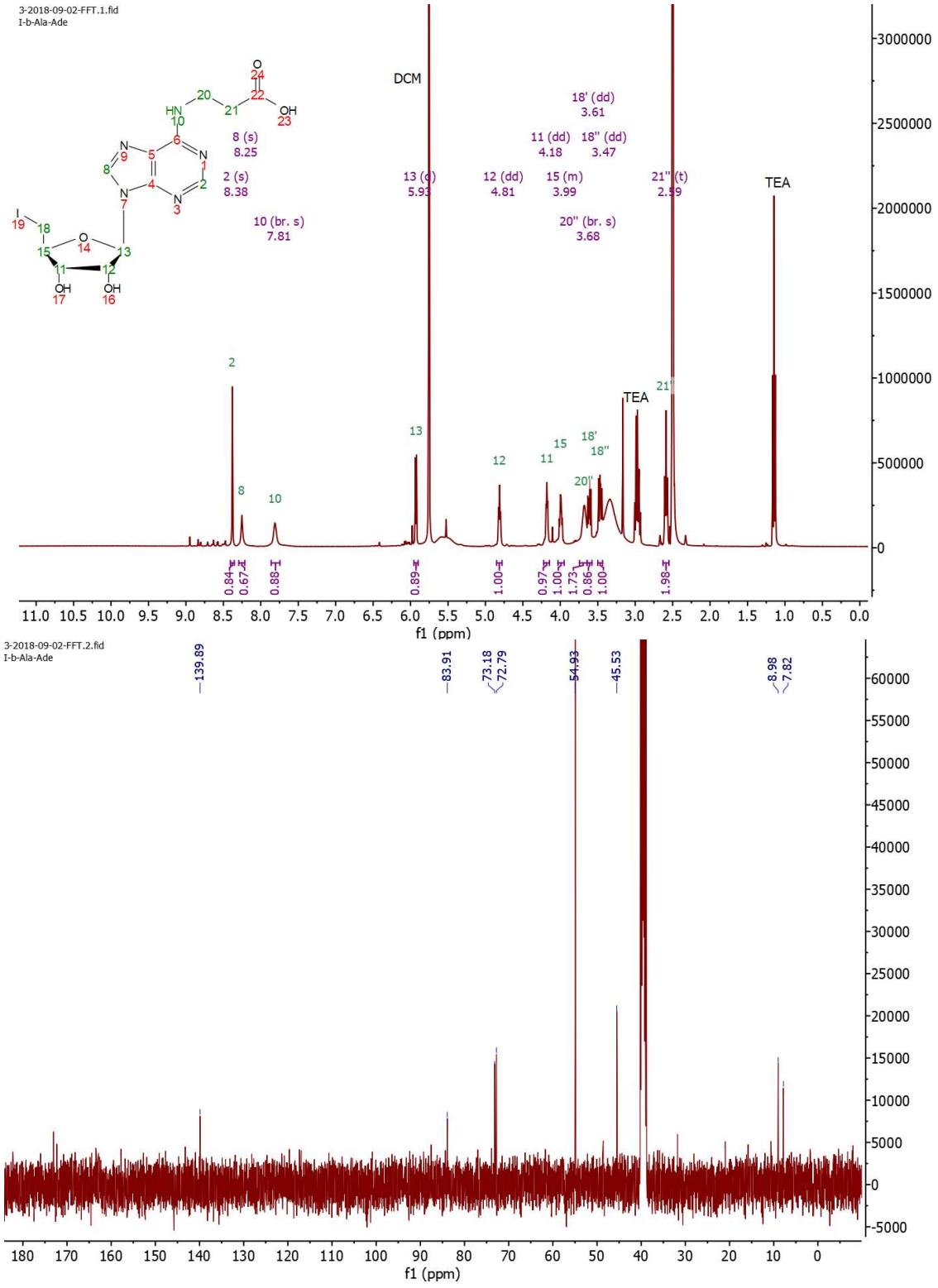
Appendix 1 Fig. 6 ^{1}H NMR (top) and ^{13}C NMR (bottom) spectra of 5’-I,6-EDA-Ade

5’-Cl,6-β-Ala-Ade



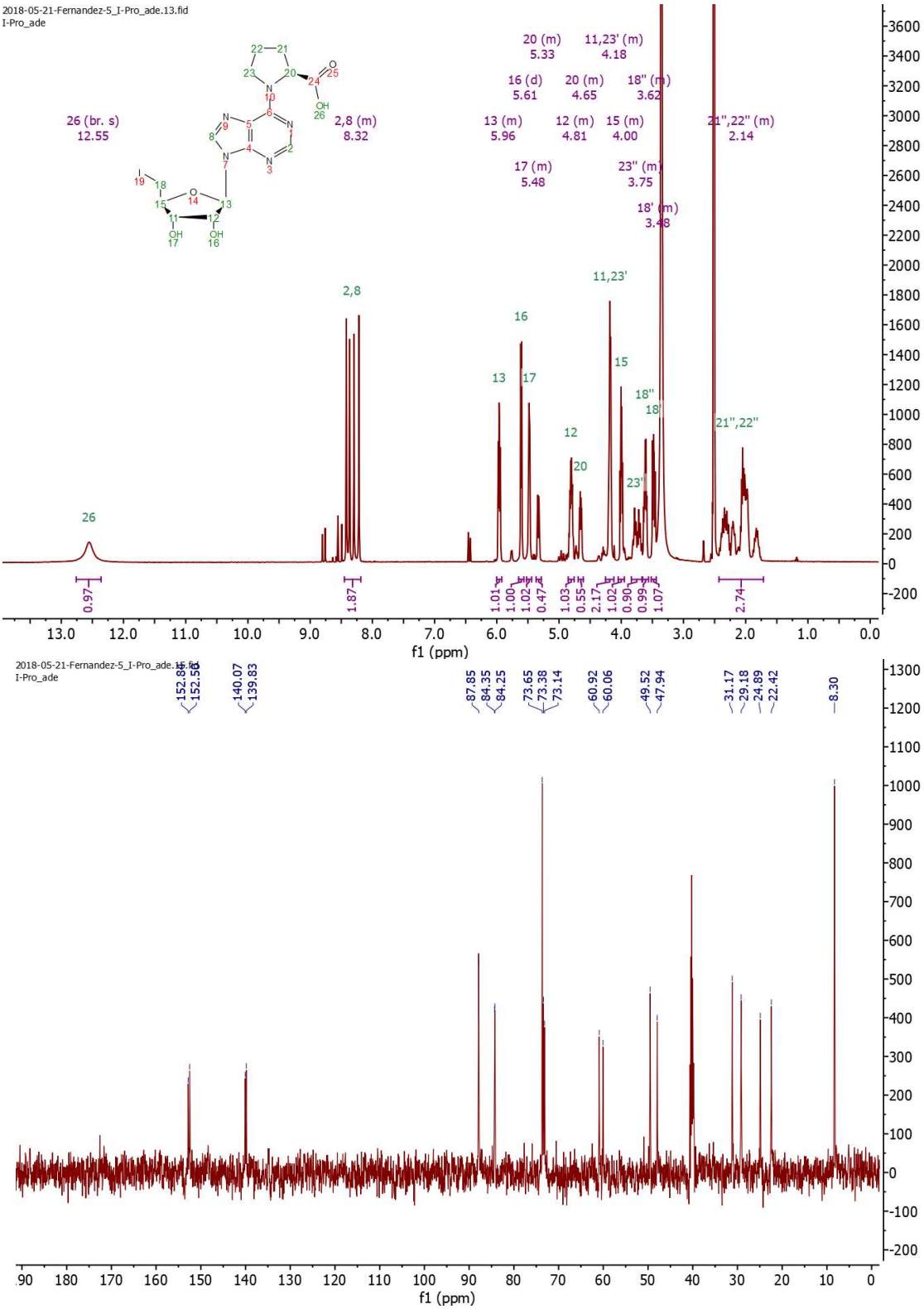
Appendix 1 Fig. 7 ^{1}H NMR (top) and ^{13}C NMR (bottom) spectra of 5’-Cl,6-β-Ala-Ade

5’-I,6-β-Ala-Ade



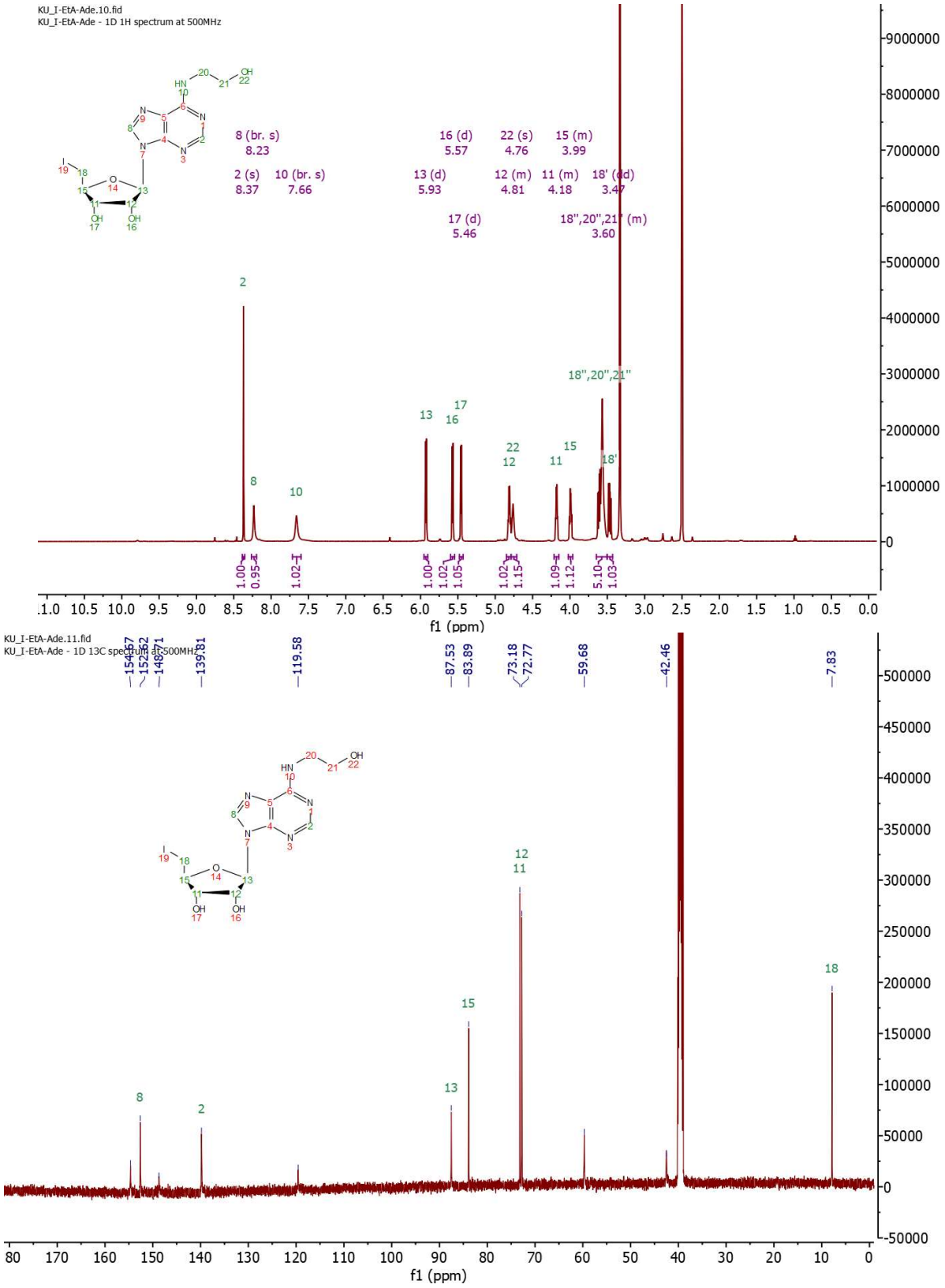
Appendix 1 Fig. 8 ^{1}H NMR (top) and ^{13}C NMR (bottom) spectra of 5’-I,6-β-Ala-Ade

5’-I,6-Pro-Ade



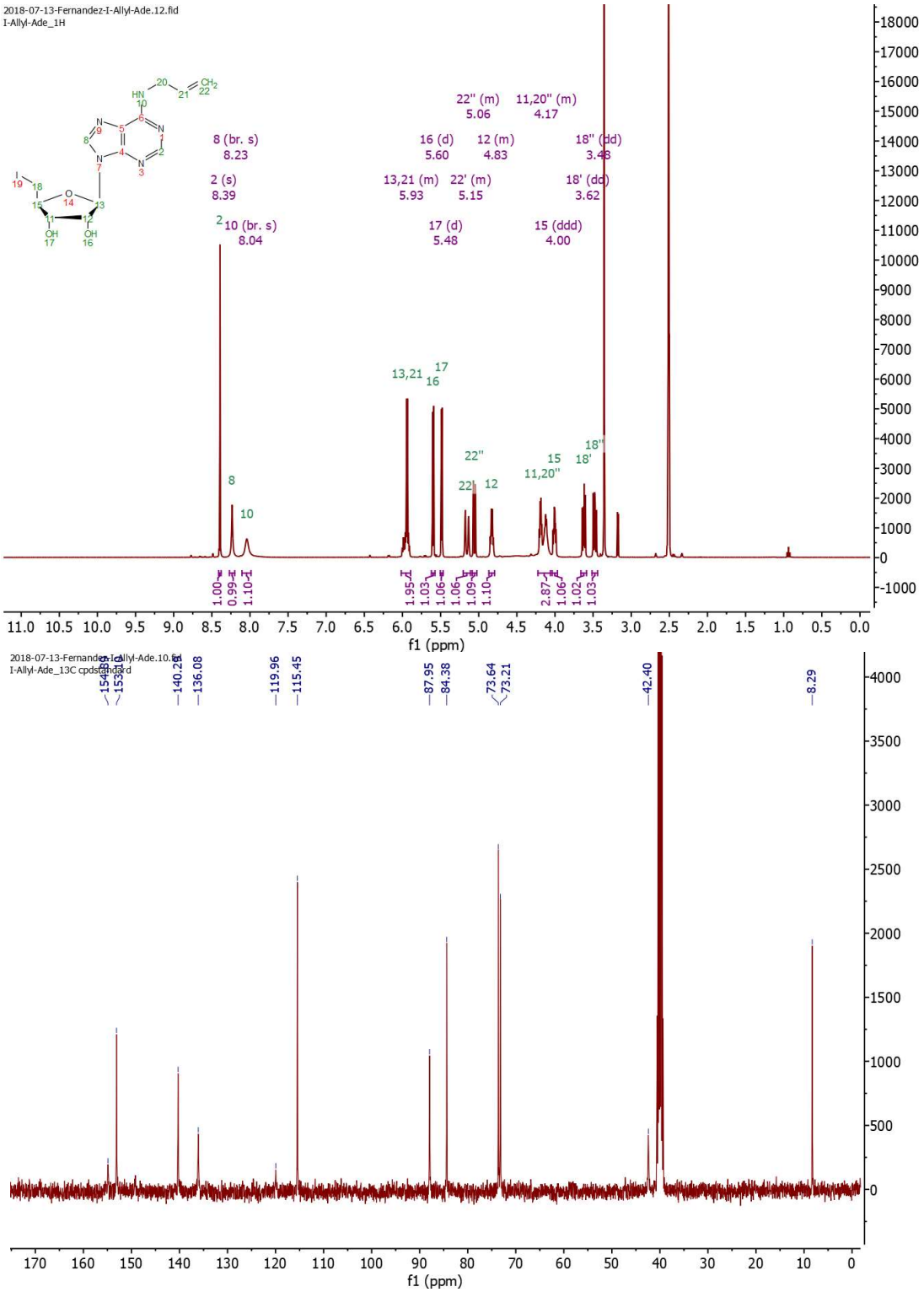
Appendix 1 Fig. 9 ^{1}H NMR (top) and ^{13}C NMR (bottom) spectra of 5’-I,6-Pro-Ade

5’-I,6-ETA-Ade



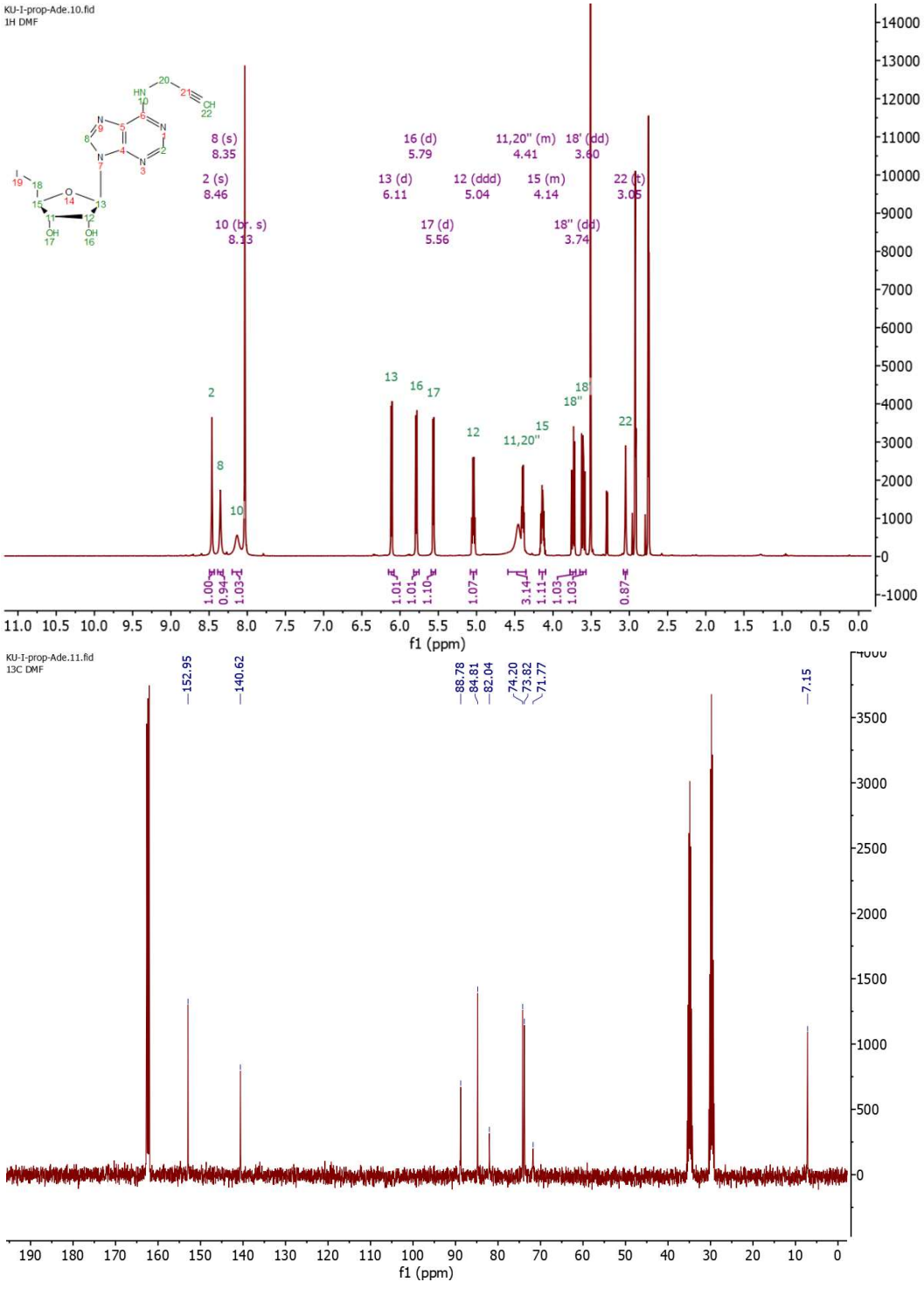
Appendix 1 Fig. 10 ^{1}H NMR (top) and ^{13}C NMR (bottom) spectra of 5’-I,6-ETA-Ade

5’-I,6-Allyl-Ade



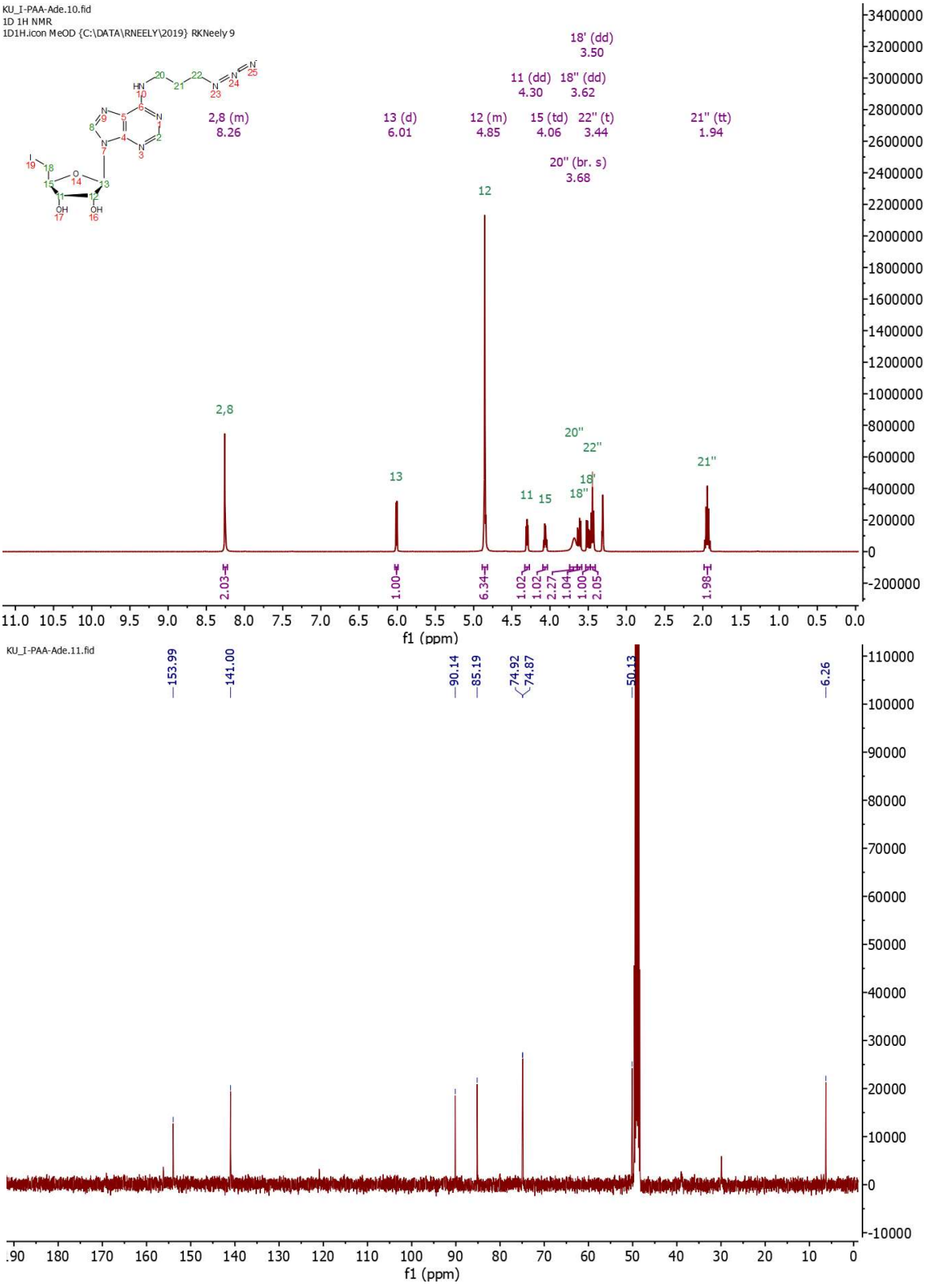
Appendix 1 Fig. 11 ^{1}H NMR (top) and ^{13}C NMR (bottom) spectra of 5’-I,6-Allyl-Ade

5’-I,6-Prop-Ade



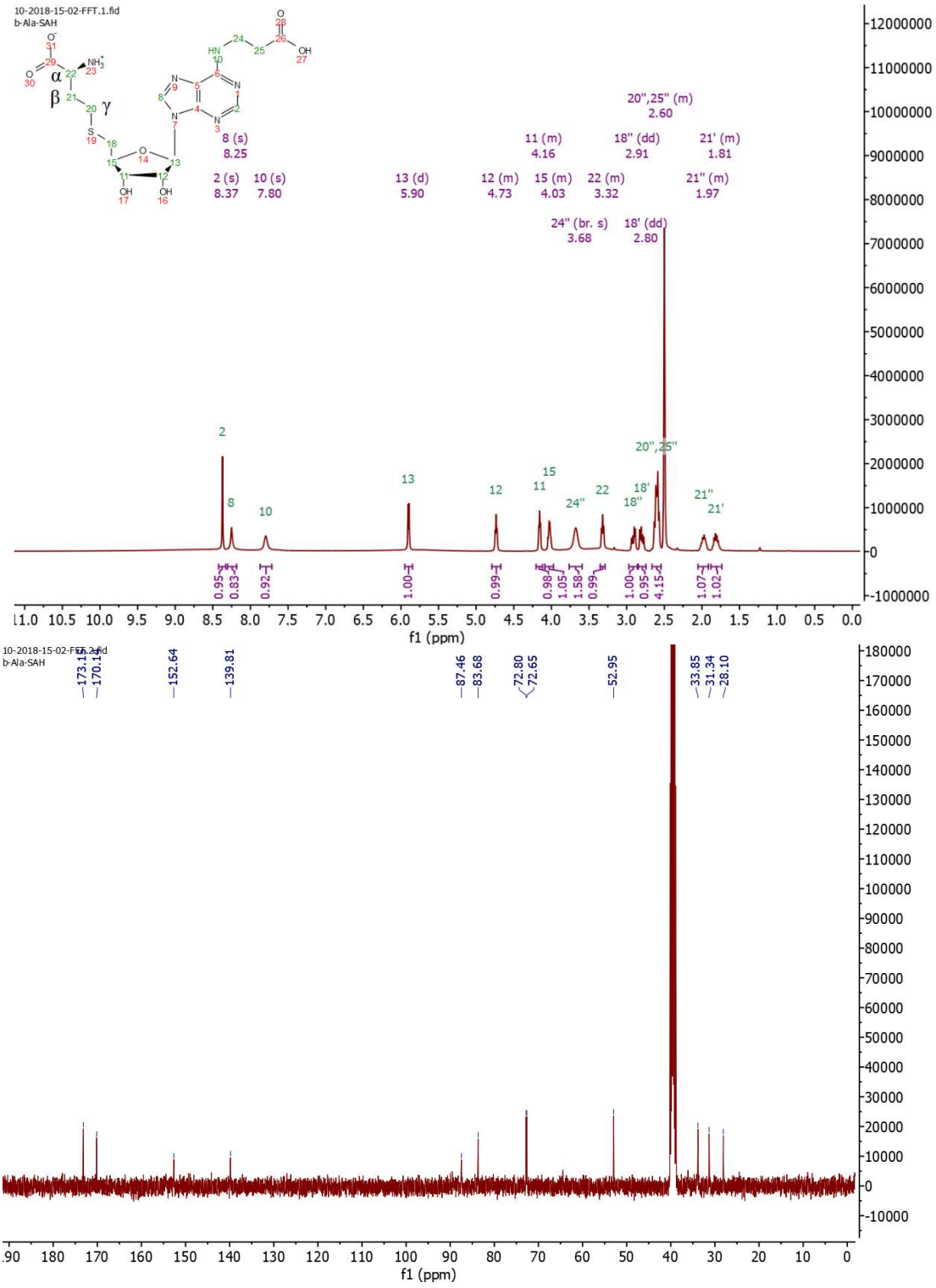
Appendix 1 Fig. 12 ^{1}H NMR (top) and ^{13}C NMR (bottom) spectra of 5’-I,6-Prop-Ade

5’-I,6-PAA-Ade



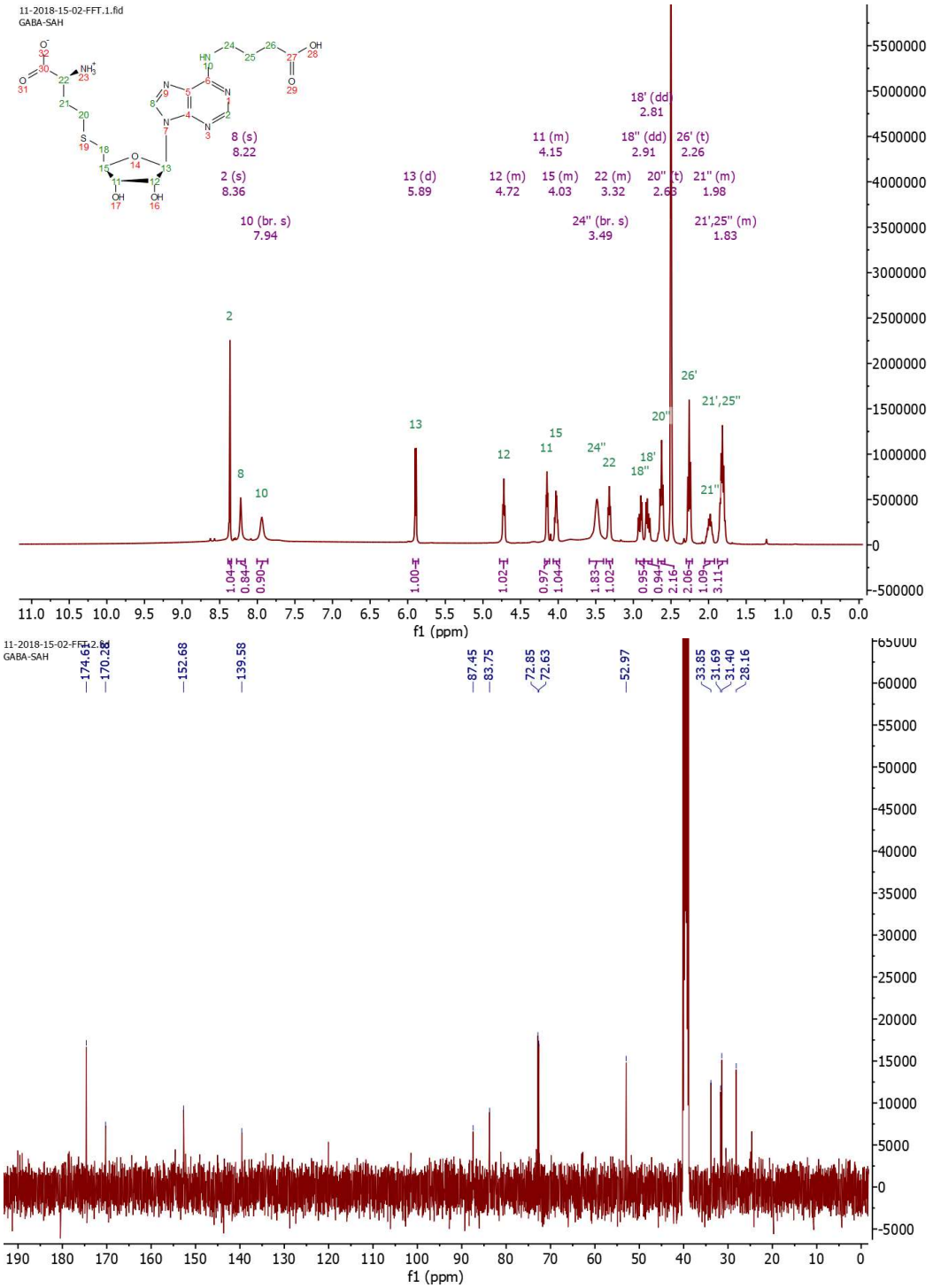
Appendix 1 Fig. 13 ^{1}H NMR (top) and ^{13}C NMR (bottom) spectra of 5’-I,6-PAA-Ade

β-Ala-AdoHcy



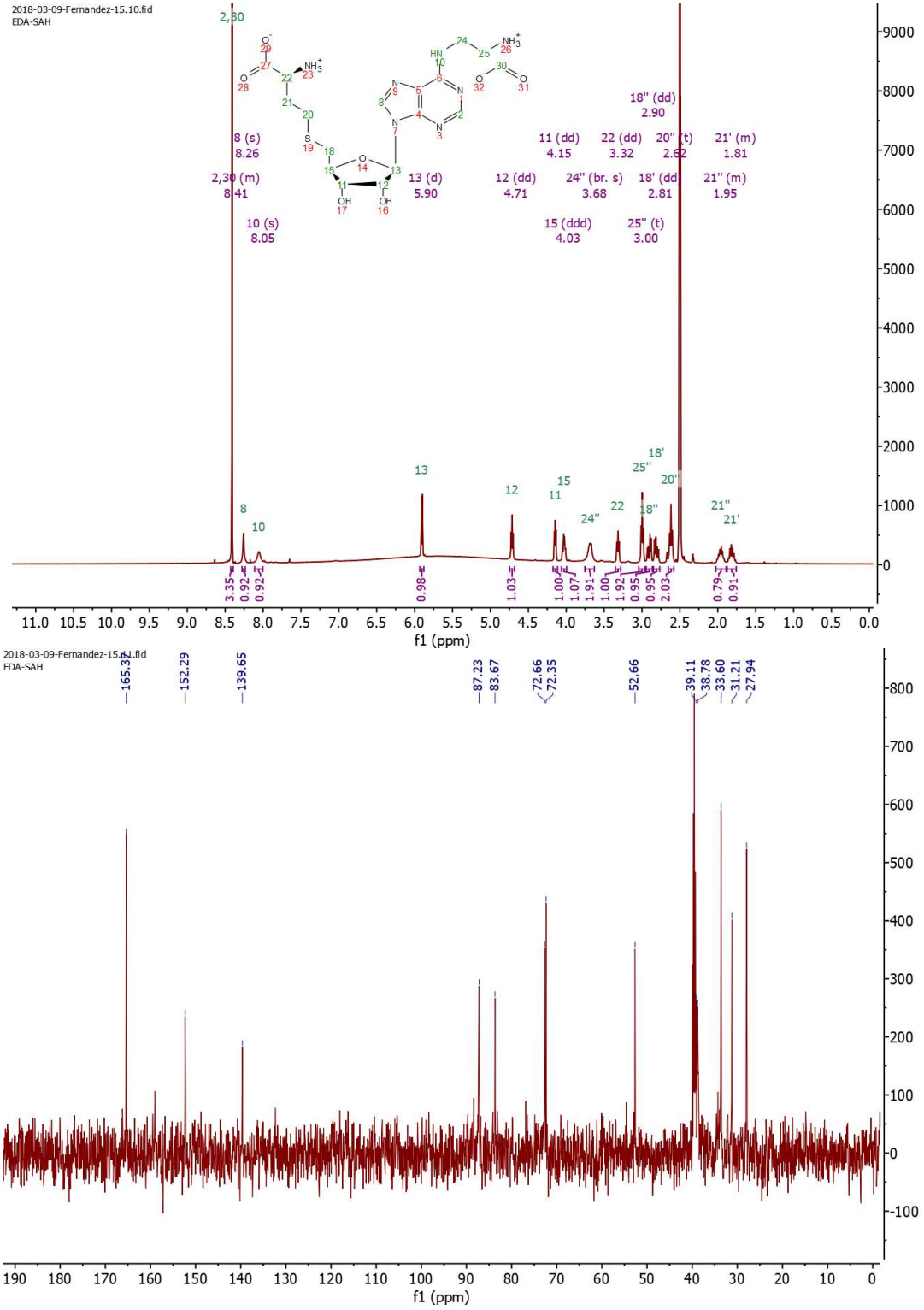
Appendix 1 Fig. 14 ^{1}H NMR (top) and ^{13}C NMR (bottom) spectra of β-Ala-AdoHcy

GABA-AdoHcy



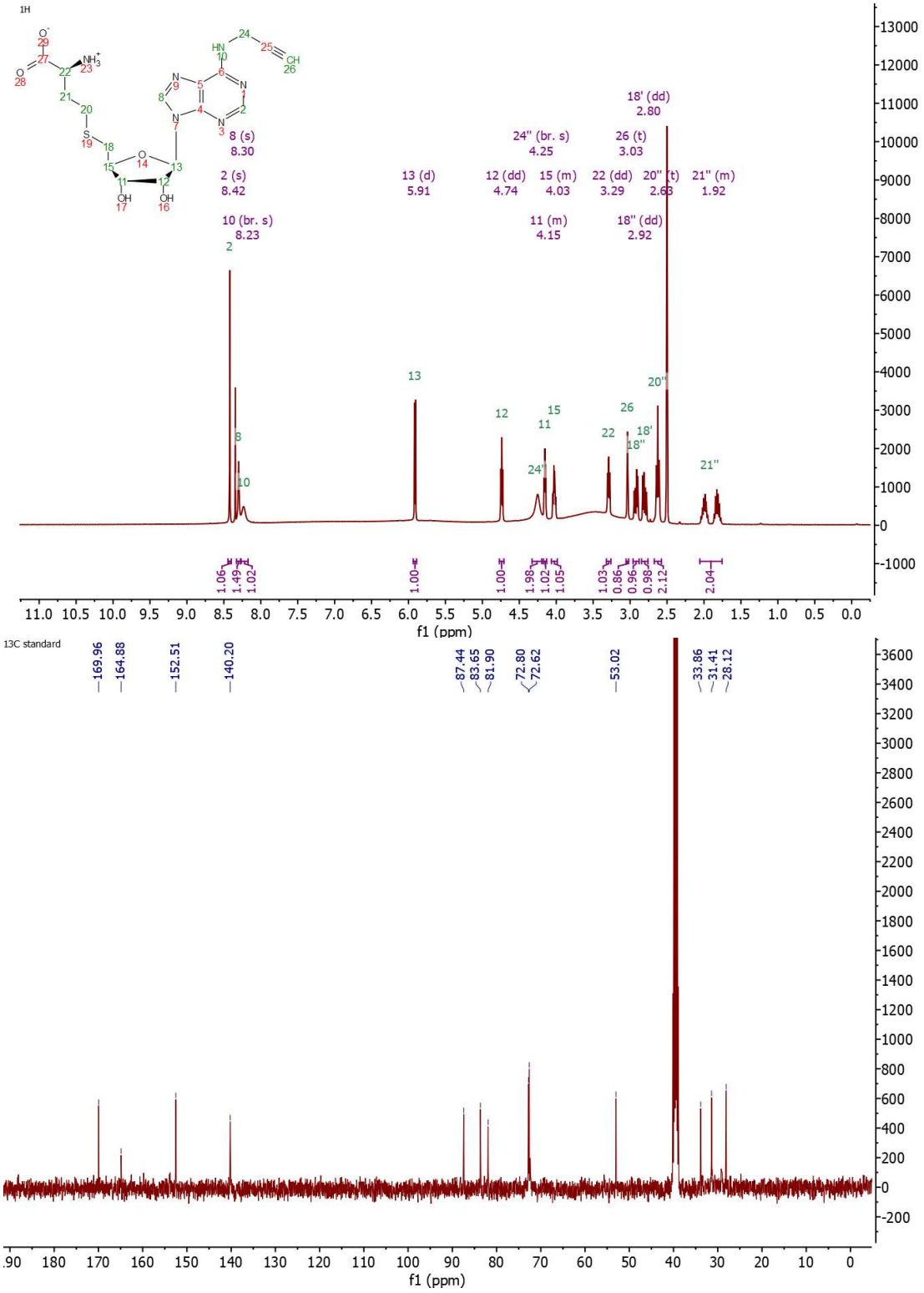
Appendix 1 Fig. 15 ^{1}H NMR (top) and ^{13}C NMR (bottom) spectra of GABA-AdoHcy

EDA-AdoHcy



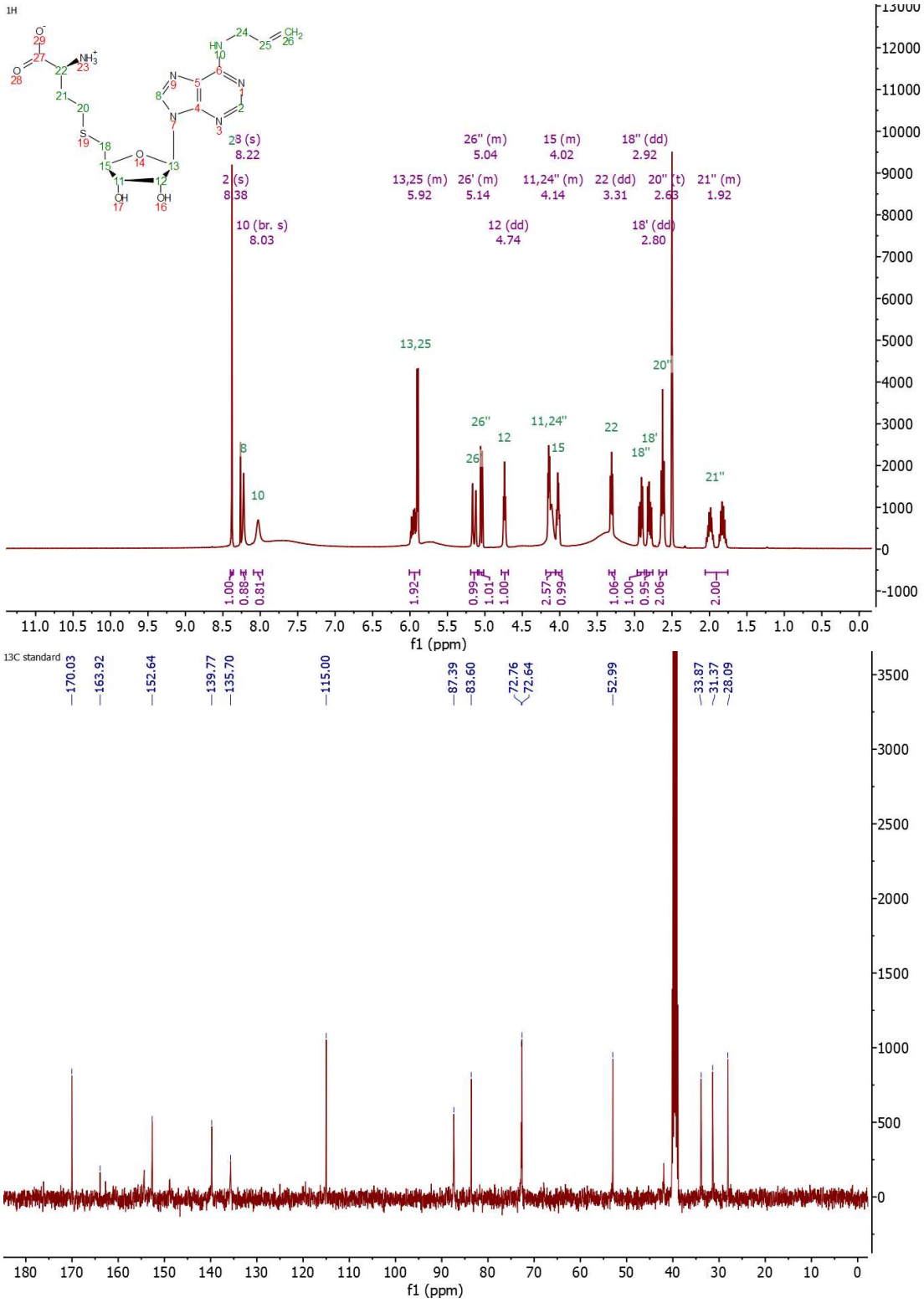
Appendix 1 Fig. 16 ^{1}H NMR (top) and ^{13}C NMR (bottom) spectra of EDA-AdoHcy

Prop-AdoHcy



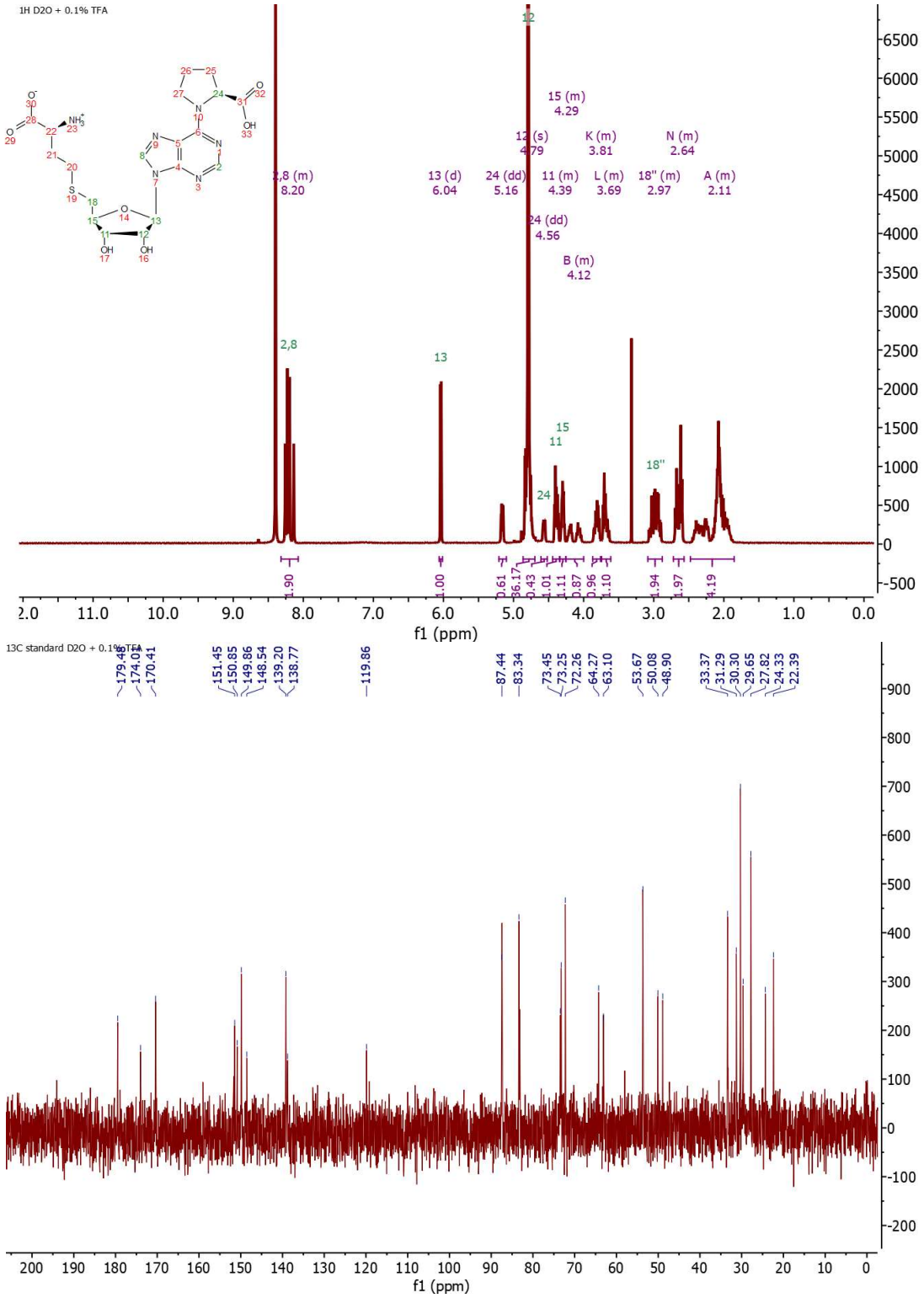
Appendix 1 Fig. 17 ^{1}H NMR (top) and ^{13}C NMR (bottom) spectra of Prop-AdoHcy

Allyl-AdoHcy



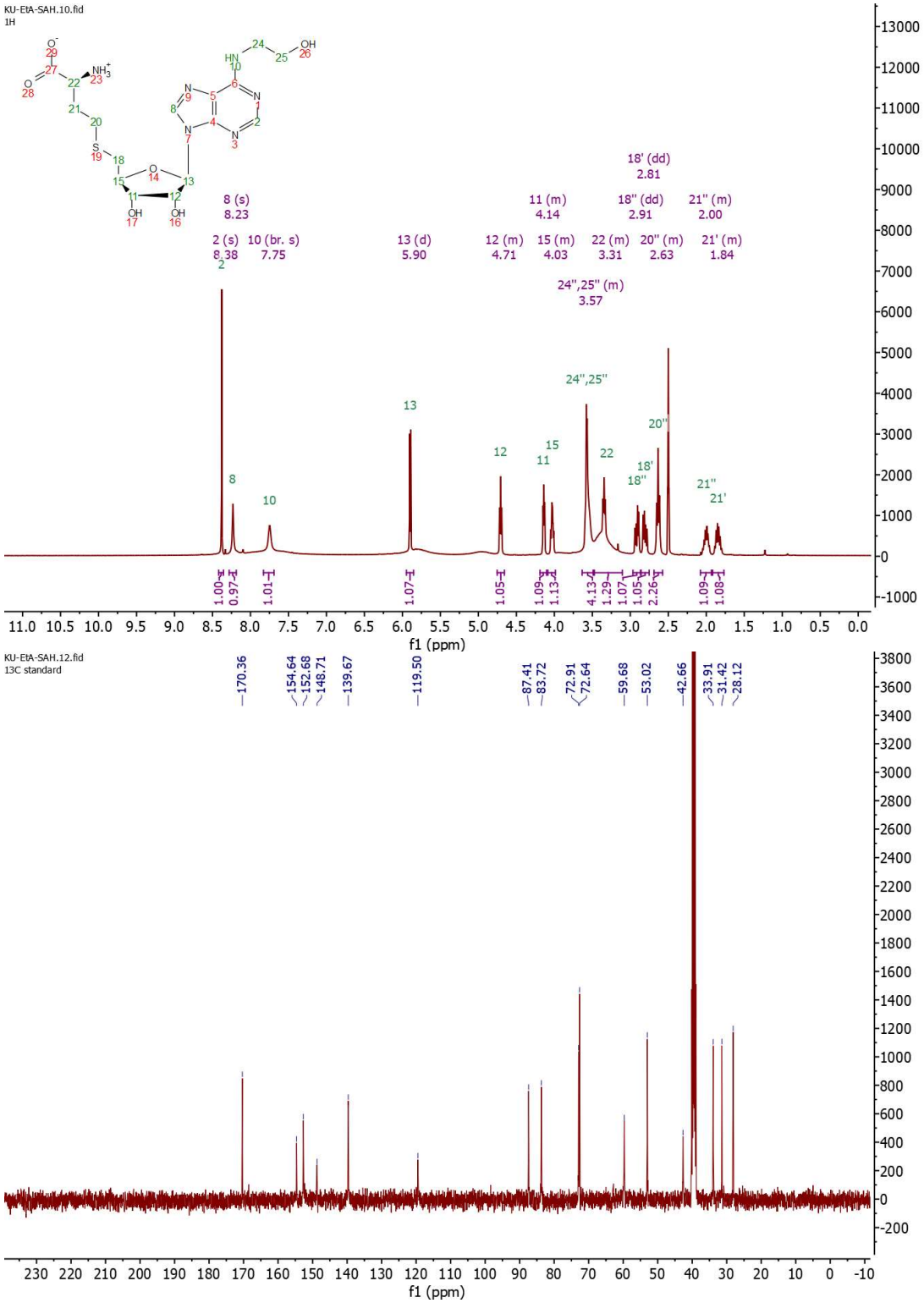
Appendix 1 Fig. 18 ^{1}H NMR (top) and ^{13}C NMR (bottom) spectra for Allyl-AdoHcy

Pro-AdoHcy



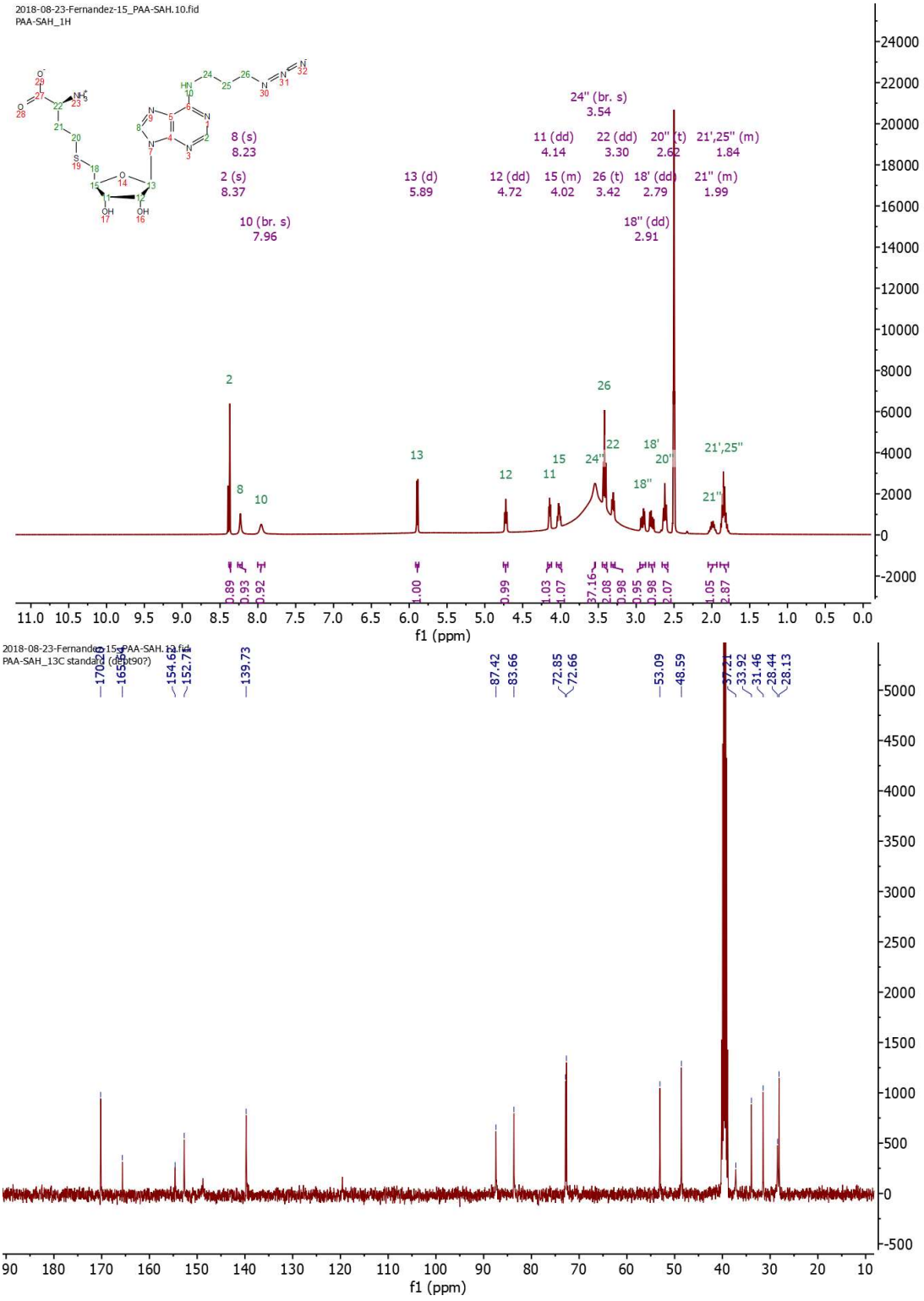
Appendix 1 Fig. 19 ^{1}H NMR (top) and ^{13}C NMR (bottom) spectra of Pro-AdoHcy

ETA-AdoHcy



Appendix 1 Fig. 20 ^{1}H NMR (top) and ^{13}C NMR (bottom) spectra of ETA-AdoHcy

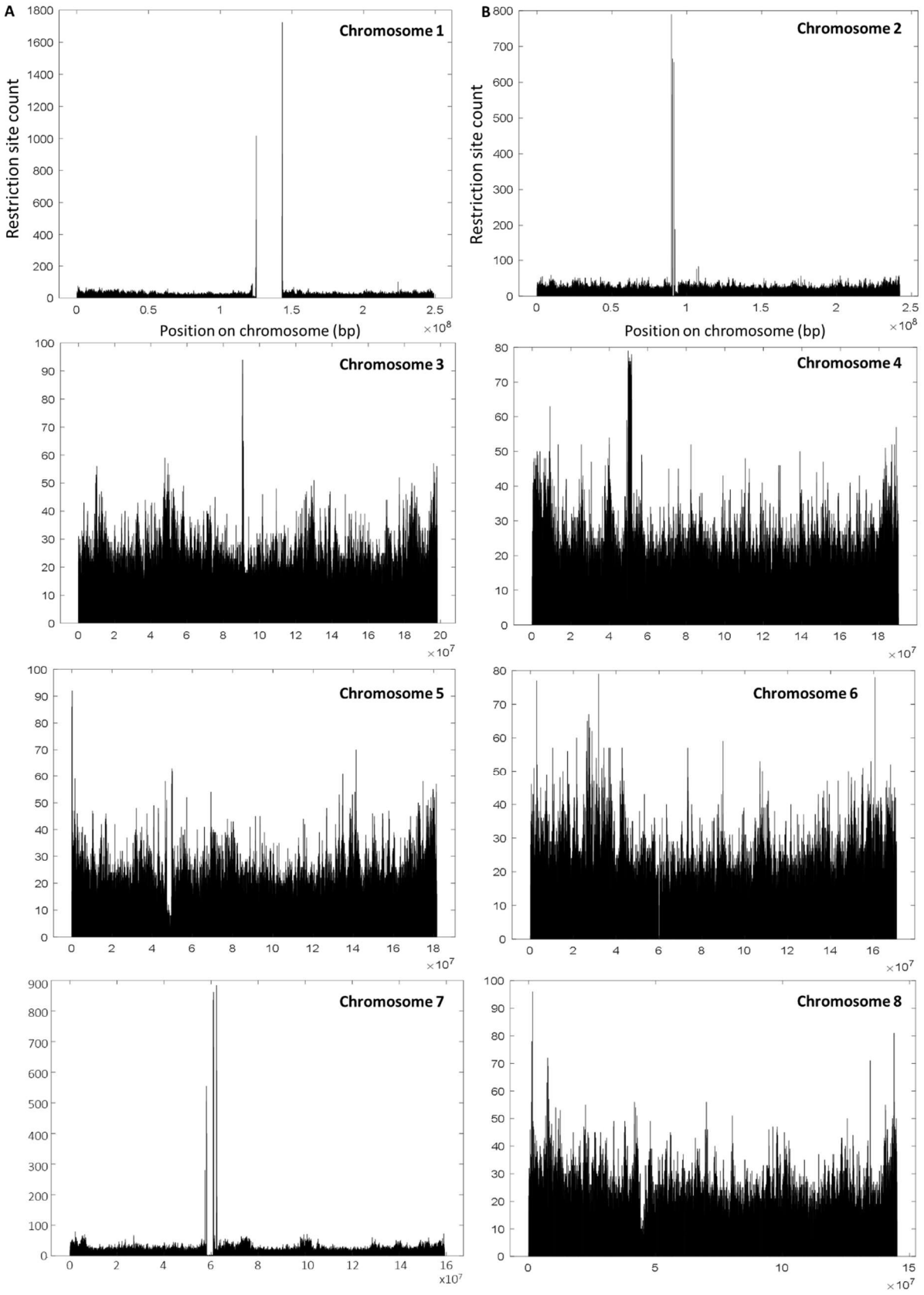
PAA-AdoHcy

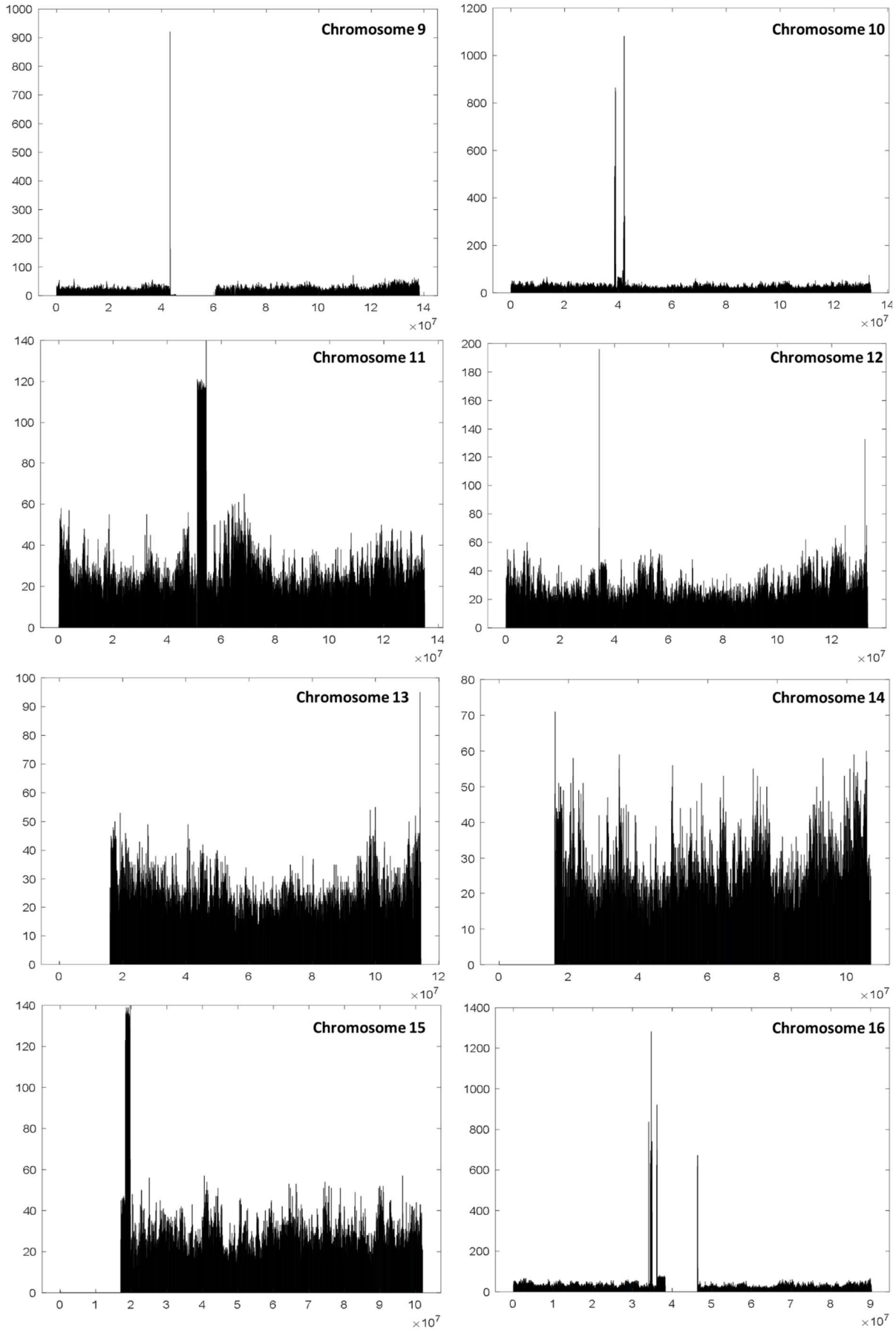


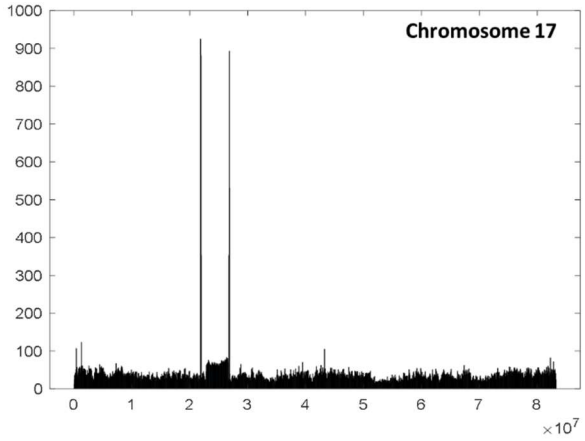
Appendix 1 Fig. 21 ^{1}H NMR (top) and ^{13}C NMR (bottom) spectra of PAA-AdoHcy

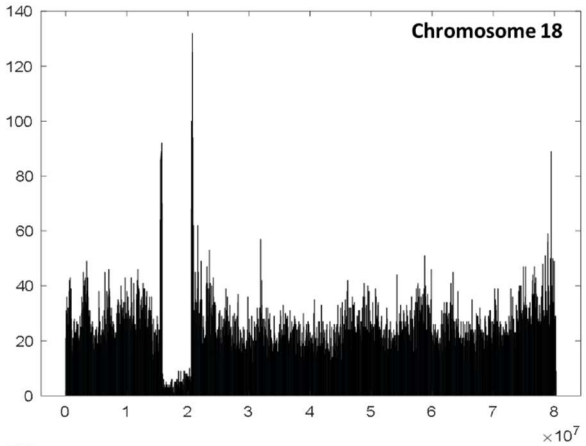
V. Appendix B

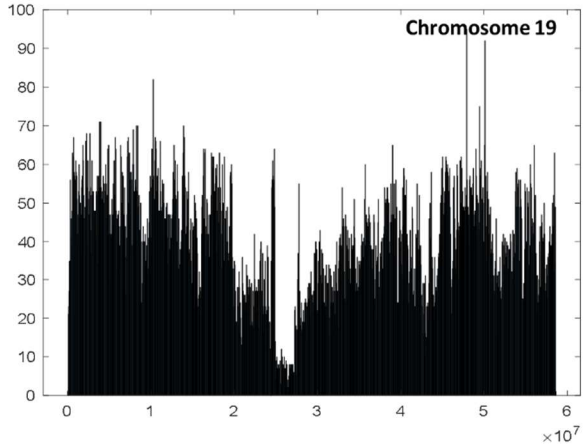
A theoretical distribution of M.TaqI target sites in each chromosome of human genome

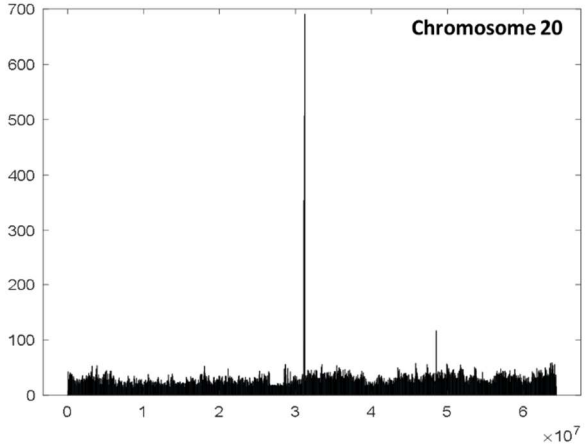

A
B
Chromosome 1
Chromosome 2
Chromosome 3
Chromosome 4
Chromosome 5
Chromosome 6
Chromosome 7
Chromosome 8
Restriction site count
Position on chromosome (bp)
×10^8
×10^7

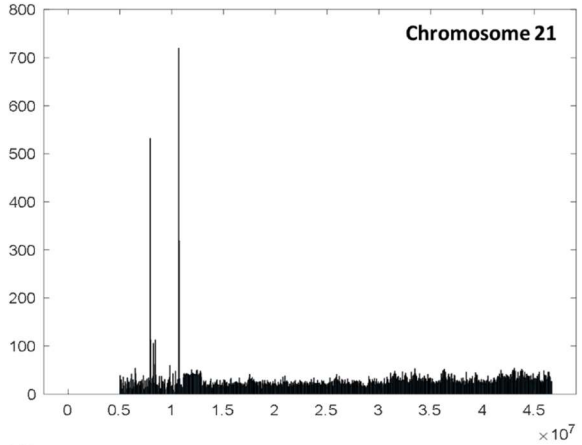

Chromosome 9
Chromosome 10
Chromosome 11
Chromosome 12
Chromosome 13
Chromosome 14
Chromosome 15
Chromosome 16
×10^7

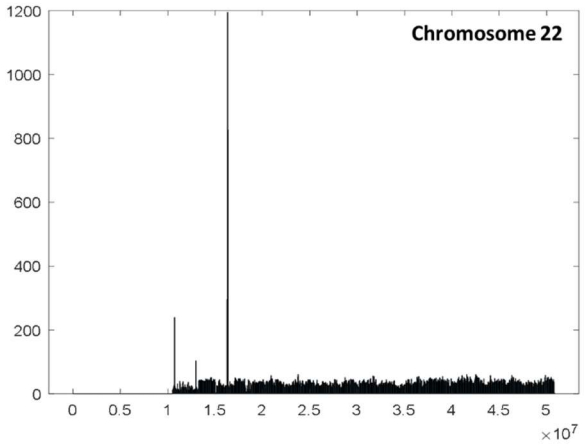

Chromosome 17
×10^7

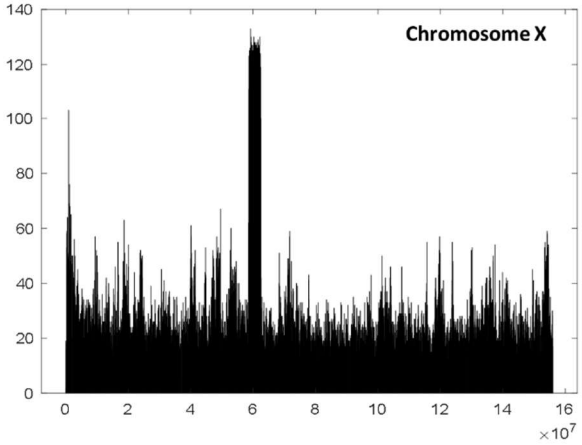

Chromosome 18
×10^7

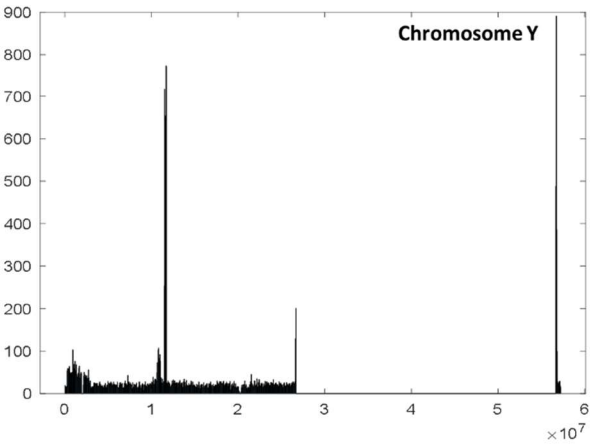

Chromosome 19
×10^7


Chromosome 20
×10^7

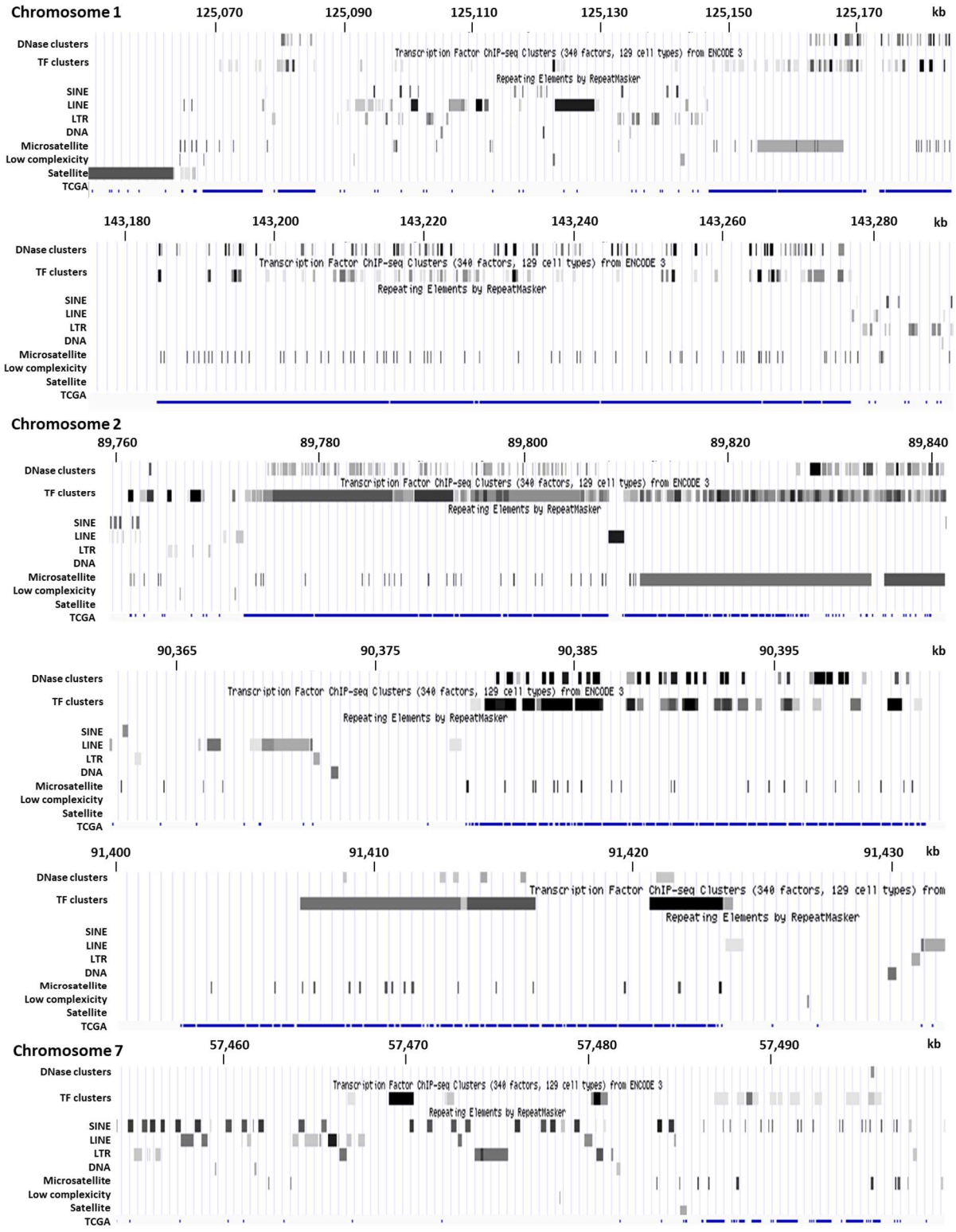

Chromosome 21
×10^7

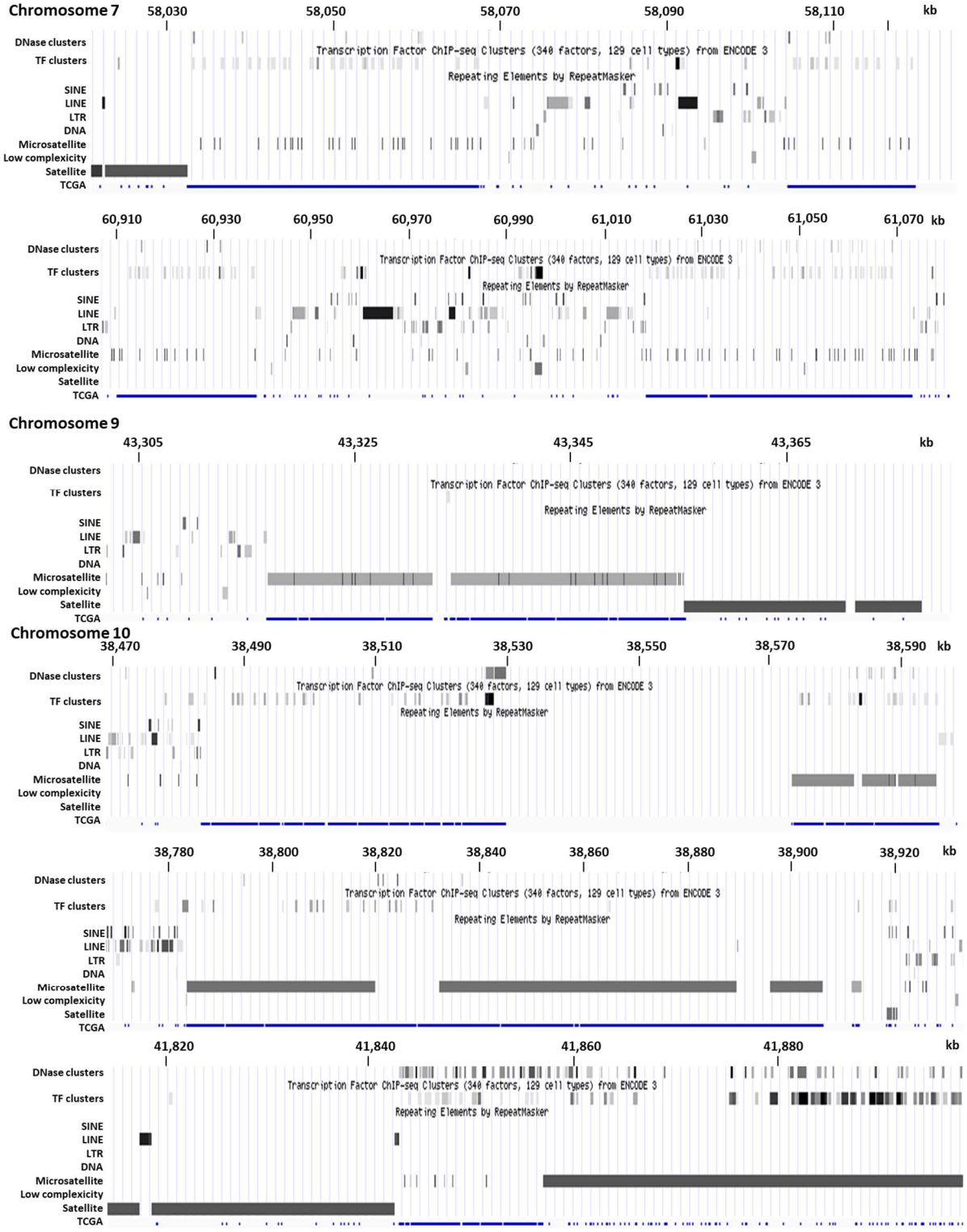

Chromosome 22
×10^7

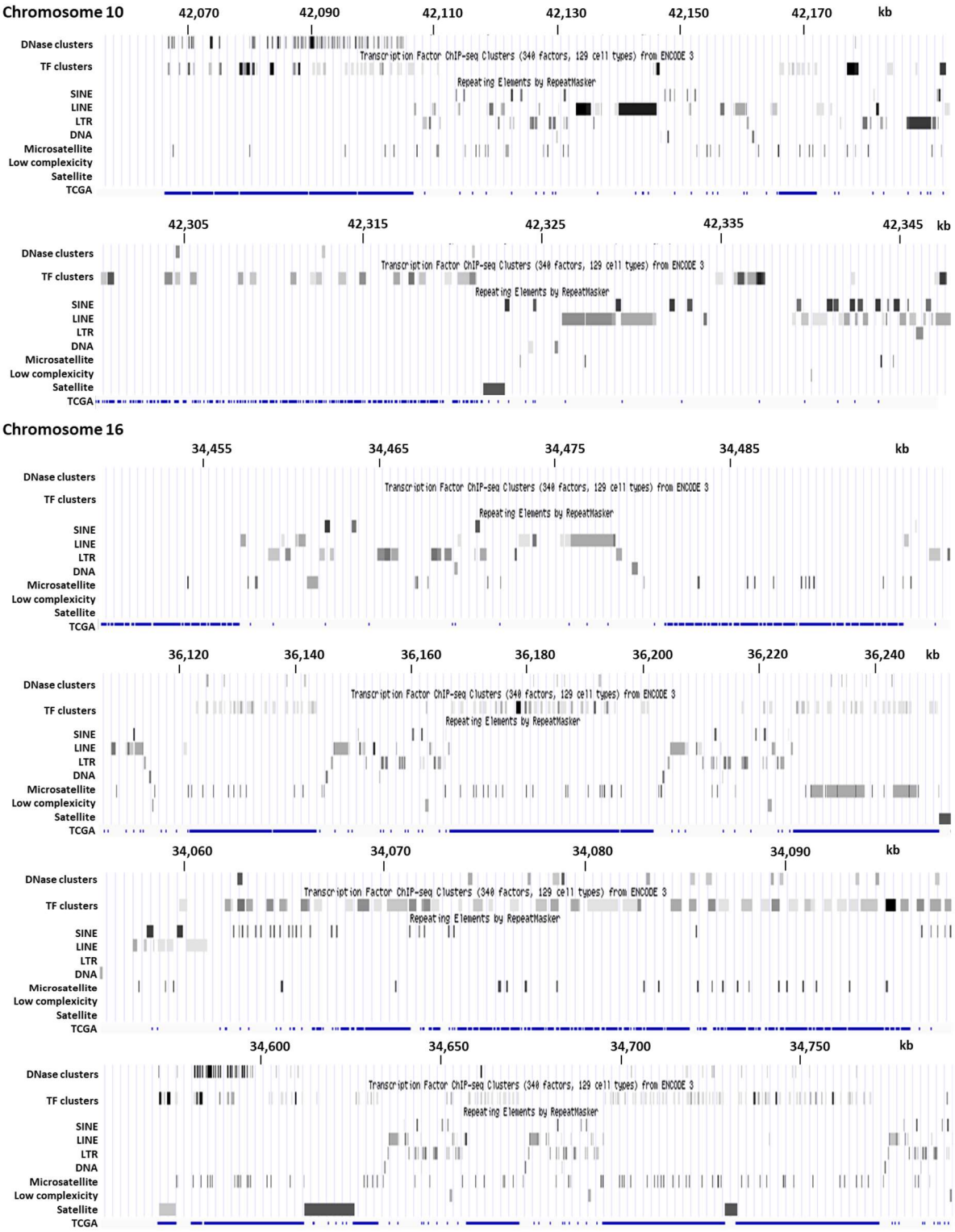

Chromosome X
×10^7

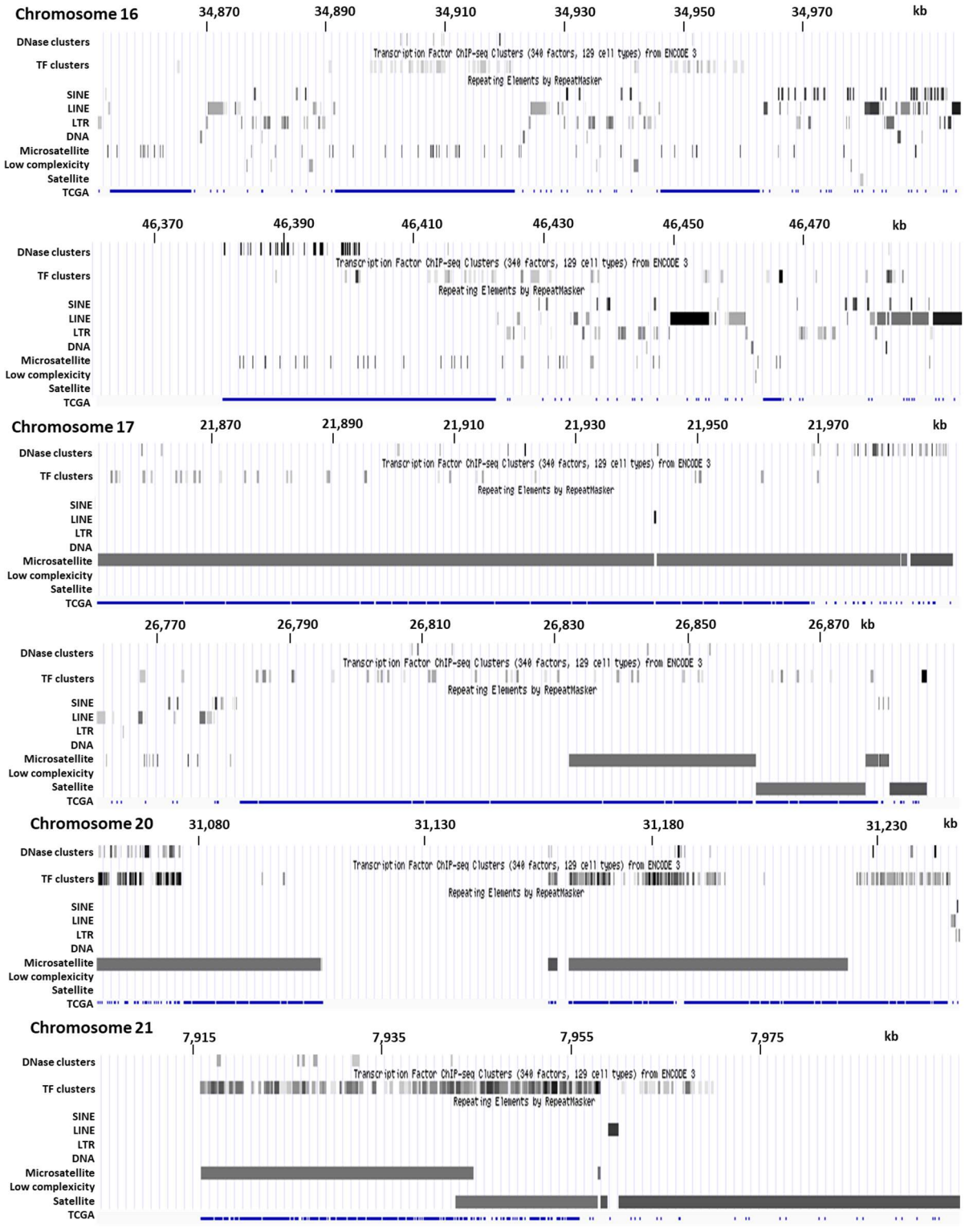

Chromosome Y
×10^7

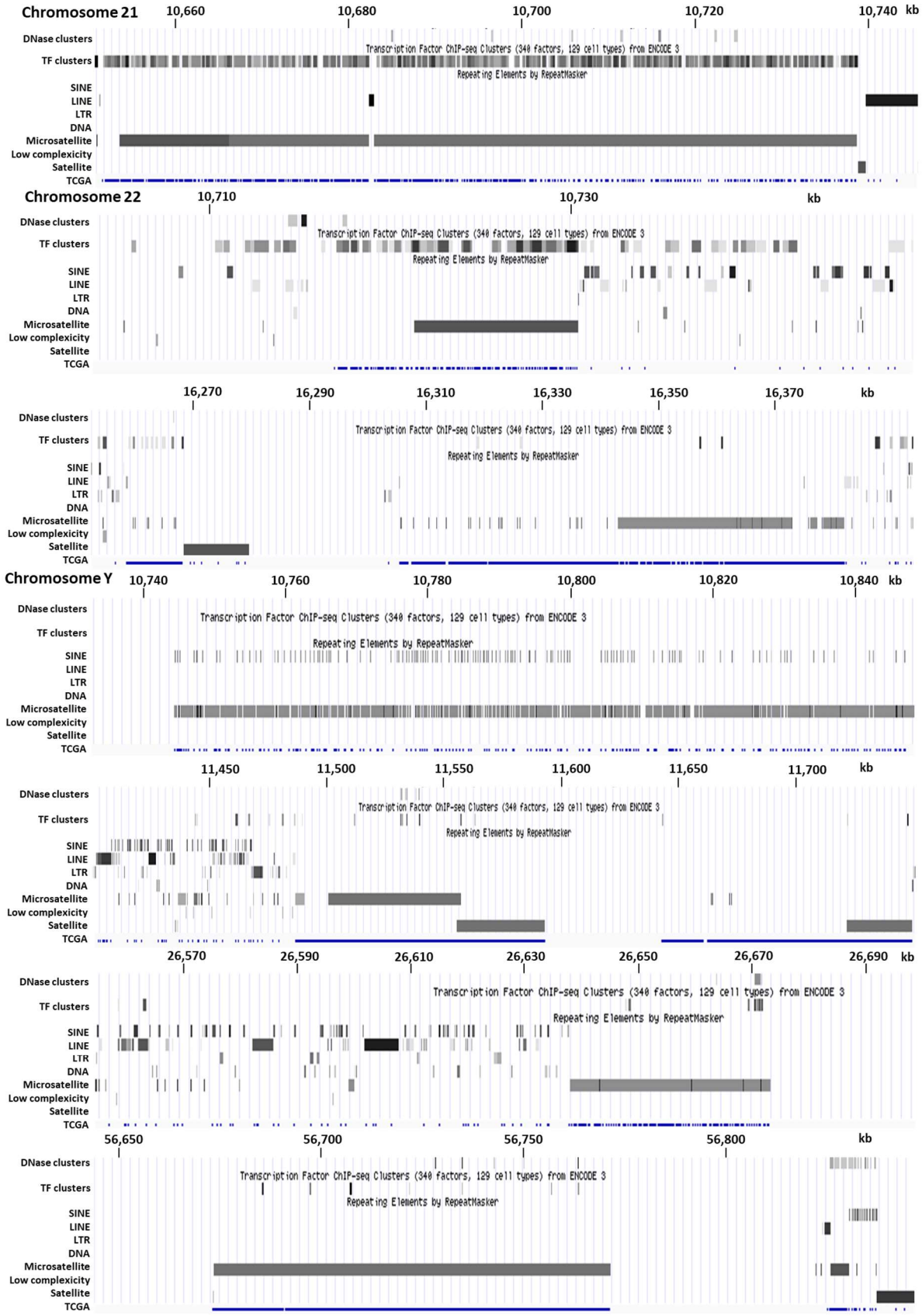
A colocalization of DNase clusters, TF clusters and repetitive sequences in the identified regions of increased density of M.TaqI target sites




Chromosome 7
58,030
58,050
58,070
58,090
58,110
kb
DNase clusters
TF clusters
Transcription Factor ChIP-seq Clusters (340 factors, 129 cell types) from ENCODE 3
Repeating Elements by RepeatMasker
SINE
LINE
LTR
DNA
Microsatellite
Low complexicity
Satellite
TCGA
60,910
60,930
60,950
60,970
60,990
61,010
61,030
61,050
61,070
kb
Chromosome 9
43,305
43,325
43,345
43,365
kb
Chromosome 10
38,470
38,490
38,510
38,530
38,550
38,570
38,590
kb
38,780
38,800
38,820
38,840
38,860
38,880
38,900
38,920
kb
41,820
41,840
41,860
41,880
kb


Chromosome 10
42,070
42,090
42,110
42,130
42,150
42,170
kb
DNase clusters
TF clusters
Transcription Factor ChIP-seq Clusters (340 factors, 129 cell types) from ENCODE 3
Repeating Elements by RepeatMasker
SINE
LINE
LTR
DNA
Microsatellite
Low complexicity
Satellite
TCGA
42,305
42,315
42,325
42,335
42,345
kb
Chromosome 16
34,455
34,465
34,475
34,485
36,120
36,140
36,160
36,180
36,200
36,220
36,240
34,060
34,070
34,080
34,090
34,600
34,650
34,700
34,750


Chromosome 16
34,870
34,890
34,910
34,930
34,950
34,970
kb
DNase clusters
TF clusters
Transcription Factor ChIP-seq Clusters (340 factors, 129 cell types) from ENCODE 3
Repeating Elements by RepeatMasker
SINE
LINE
LTR
DNA
Microsatellite
Low complexicity
Satellite
TCGA
46,370
46,390
46,410
46,430
46,450
46,470
kb
Chromosome 17
21,870
21,890
21,910
21,930
21,950
21,970
kb
26,770
26,790
26,810
26,830
26,850
26,870
kb
Chromosome 20
31,080
31,130
31,180
31,230
kb
Chromosome 21
7,915
7,935
7,955
7,975
kb


Chromosome 21
10,660
10,680
10,700
10,720
10,740 kb
DNase clusters
TF clusters
Transcription Factor ChIP-seq Clusters (340 factors, 129 cell types) from ENCODE 3
Repeating Elements by RepeatMasker
SINE
LINE
LTR
DNA
Microsatellite
Low complexicity
Satellite
TCGA
Chromosome 22
10,710
10,730
kb
16,270
16,290
16,310
16,330
16,350
16,370
kb
Chromosome Y
10,740
10,760
10,780
10,800
10,820
10,840 kb
11,450
11,500
11,550
11,600
11,650
11,700 kb
26,570
26,590
26,610
26,630
26,650
26,670
26,690 kb
56,650
56,700
56,750
56,800
kb

Fiji macro script used for deconvolution of ExM images

```
name = File.nameWithoutExtension;path = File.directory;
rename("stack_1");
selectWindow("stack_1");
run("Split Channels");
selectWindow("C1-stack_1");
Stack.getStatistics(voxelCount, mean, min, max, stdDev);
setMinAndMax(min, max);
run("Iterative Deconvolve 3D", "image=C1-stack_1 point=PSF405.tif "
+ "output=C1-stack_1_decon normalize show log perform wiener=0.0 "
+ "low=1 z_direction1 maximum=100 terminate=0.01");
resetMinAndMax();
saveAs(".tif", path+"405_"+name+"_decon");
C1_decon = getTitle();
run("Duplicate...", "duplicate");
run("Gaussian Blur...", "sigma=1 stack");
saveAs(".tif", path+"405_"+name+"_decon_blur");
C1_decon_blur = getTitle();
selectWindow("C2-stack_1");
Stack.getStatistics(voxelCount, mean, min, max, stdDev);
setMinAndMax(min, max);
run("Iterative Deconvolve 3D", "image=C2-stack_1 point=PSF488.tif "
+ "output=C2-stack_1_decon normalize show log perform wiener=0.0 "
+ "low=1 z_direction=1 maximum=100 terminate=0.01");
resetMinAndMax();
saveAs(".tif", path+"488_"+name+"_decon");
C2_decon = getTitle();
run("Duplicate...", "duplicate");
```

```
run("Gaussian Blur...", "sigma=1 stack");

saveAs(".tif", path+"488_"+name+"_decon_blur");

C2_decon_blur = getTitle();

run("Merge Channels...", "blue=[" + C1_decon + "] green=[" + C2_decon + "] create
keep");

saveAs(".tif", path+name+"_decon_merged");

run("Merge Channels...", "blue=[" + C1_decon_blur + "] green=[" + C2_decon_blur +
"] create keep");

saveAs(".tif", path+name+"_decon_blur_merged");
```

Fiji macro script used for 3D segmentation of ExM images

```
//open segmented file first and set voxel size prior the analysis

name = File.nameWithoutExtension;

name_segmentation = name + "_Simple Segmentation__"

path = File.directory;

selectWindow(name + ".tif");

Stack.getStatistics(voxelCount, mean, min, max, stdDev);

setMinAndMax(min, max);

run("Subtract Background...", "rolling=50 stack");

run("Duplicate...", "duplicate");

saveAs(".tif", path + name + "_substracted");

run("Close");

selectWindow(name_segmentation + ".tiff");

run("3D Manager Options", "volume surface compactness integrated_density
mean_grey_value std_dev_grey_value minimum_grey_value maximum_grey_value
centroid_(pix) centroid_(unit) distance_to_surface centre_of_mass_(pix)
centre_of_mass_(unit) radial_distance surface_contact
exclude_objects_on_edges_xy exclude_objects_on_edges_z
distance_between_centers=10 distance_max_contact=1.80 drawing=Contour");

run("3D Manager");
```

```
Ext.Manager3D_Segment(255, 255);

Ext.Manager3D_AddImage();

Ext.Manager3D_Save(path + name + "_3D_ROI" + ".zip");

selectWindow(name + ".tif");

Ext.Manager3D_SelectAll();

Ext.Manager3D_Measure();

Ext.Manager3D_SaveResult("M",path + name +"_3D_M" + ".csv");

Ext.Manager3D_CloseResult("M");

// Use Q for the Quantification window, D for distances, C for colocalisation, L or V for list voxels and A for A windows

// if list is not visible please refresh list by using Deselect

Ext.Manager3D_Quantif();

Ext.Manager3D_SaveResult("Q",path + name + "_3D_Q" + ".csv");

Ext.Manager3D_CloseResult("Q");

run("Close All")
```

References

1. Rosenthal, N., DNA and the Genetic Code. *The New England Journal of Medicine* **2010,** *331*, 39-41.
2. Venter, J. C.; Adams, M. D.; Myers, E. W.; Li, P. W.; Mural, R. J.; Sutton, G. G.; Smith, H. O.; Yandell, M.; Evans, C. A.; Holt, R. A.; Gocayne, J. D.; Amanatides, P.; Ballew, R. M.; Huson, D. H.; Wortman, J. R.; Zhang, Q.; Kodira, C. D.; Zheng, X. H.; Chen, L.; Skupski, M.; Subramanian, G.; Thomas, P. D.; Zhang, J.; Miklos, G. L. G.; Nelson, C.; Broder, S.; Clark, A. G.; Nadeau, J.; McKusick, V. A.; Zinder, N.; Levine, A. J.; Roberts, R. J.; Simon, M.; Slayman, C.; Hunkapiller, M.; Bolanos, R.; Delcher, A.; Dew, I.; Fasulo, D.; Flanigan, M.; Florea, L.; Halpern, A.; Hannenhalli, S.; Kravitz, S.; Levy, S.; Mobarry, C.; Reinert, K.; Remington, K.; Abu-Threideh, J.; Beasley, E.; Biddick, K.; Bonazzi, V.; Brandon, R.; Cargill, M.; Chandramouliswaran, I.; Charlab, R.; Chaturvedi, K.; Deng, Z.; Francesco, V. D.; Dunn, P.; Eilbeck, K.; Evangelista, C.; Gabrielian, A. E.; Gan, W.; Ge, W.; Gong, F.; Gu, Z.; Guan, P.; Heiman, T. J.; Higgins, M. E.; Ji, R.-R.; Ke, Z.; Ketchum, K. A.; Lai, Z.; Lei, Y.; Li, Z.; Li, J.; Liang, Y.; Lin, X.; Lu, F.; Merkulov, G. V.; Milshina, N.; Moore, H. M.; Naik, A. K.; Narayan, V. A.; Neelam, B.; Nusskern, D.; Rusch, D. B.; Salzberg, S.; Shao, W.; Shue, B.; Sun, J.; Wang, Z. Y.; Wang, A.; Wang, X.; Wang, J.; Wei, M.-H.; Wides, R.; Xiao, C.; Yan, C.; Yao, A.; Ye, J.; Zhan, M.; Zhang, W.; Zhang, H.; Zhao, Q.; Zheng, L.; Zhong, F.; Zhong, W.; Zhu, S. C.; Zhao, S.; Gilbert, D.; Baumhueter, S.; Spier, G.; Carter, C.; Cravchik, A.; Woodage, T.; Ali, F.; An, H.; Awe, A.; Baldwin, D.; Baden, H.; Barnstead, M.; Barrow, I.; Beeson, K.; Busam, D.; Carver, A.; Center, A.; Cheng, M. L.; Curry, L.; Danaher, S.; Davenport, L.; Desilets, R.; Dietz, S.; Dodson, K.; Doup, L.; Ferriera, S.; Garg, N.; Gluecksmann, A.; Hart, B.; Haynes, J.; Haynes, C.; Heiner, C.; Hladun, S.; Hostin, D.; Houck, J.; Howland, T.; Ibegwam, C.; Johnson, J.; Kalush, F.; Kline, L.; Koduru, S.; Love, A.; Mann, F.; May, D.; McCawley, S.; McIntosh, T.; McMullen, I.; Moy, M.; Moy, L.; Murphy, B.; Nelson, K.; Pfannkoch, C.; Pratts, E.; Puri, V.; Qureshi, H.; Reardon, M.; Rodriguez, R.; Rogers, Y.-H.; Romblad, D.; Ruhfel, B.; Scott, R.; Sitter, C.; Smallwood, M.; Stewart, E.; Strong, R.; Suh, E.; Thomas, R.; Tint, N. N.; Tse, S.; Vech, C.; Wang, G.; Wetter, J.; Williams, S.; Williams, M.; Windsor, S.; Winn-Deen, E.; Wolfe, K.; Zaveri, J.; Zaveri, K.; Abril, J. F.; Guigó, R.; Campbell, M. J.; Sjolander, K. V.; Karlak, B.; Kejariwal, A.; Mi, H.; Lazareva, B.; Hatton, T.; Narechania, A.; Diemer, K.; Muruganujan, A.; Guo, N.; Sato, S.; Bafna, V.; Istrail, S.; Lippert, R.; Schwartz, R.; Walenz, B.; Yooseph, S.; Allen, D.; Basu, A.; Baxendale, J.; Blick, L.; Caminha, M.; Carnes-Stine, J.; Caulk, P.; Chiang, Y.-H.; Coyne, M.; Dahlke, C.; Mays, A. D.; Dombroski, M.; Donnelly, M.; Ely, D.; Esparham, S.; Fosler, C.; Gire, H.; Glanowski, S.; Glasser, K.; Glodek, A.; Gorokhov, M.; Graham, K.; Gropman, B.; Harris, M.; Heil, J.; Henderson, S.; Hoover, J.; Jennings, D.; Jordan, C.; Jordan, J.; Kasha, J.; Kagan, L.; Kraft, C.; Levitsky, A.; Lewis, M.; Liu, X.; Lopez, J.; Ma, D.; Majoros, W.; McDaniel, J.; Murphy, S.; Newman, M.; Nguyen, T.; Nguyen, N.; Nodell, M.; Pan, S.; Peck, J.; Peterson, M.; Rowe, W.; Sanders, R.; Scott, J.; Simpson, M.; Smith, T.; Sprague, A.; Stockwell, T.; Turner, R.; Venter, E.; Wang, M.; Wen, M.; Wu, D.; Wu, M.; Xia, A.; Zandieh, A.; Zhu, X., The Sequence of the Human Genome. *Science* **2001,** *291* (5507), 1304-1351.
3. Rauscher, F. J., It Is Time for a Human Epigenome Project. *Cancer Research* **2005,** *65* (24), 11229.

4. Bradley E. Bernstein, A. M., Eric S. Lander, The Mammalian Epigenome. *Cell* **2007,** *128* (4), 669-681.
5. Feinberg, A. P.; Tycko, B., The history of cancer epigenetics. *Nature Reviews Cancer* **2019,** *4* (2), 143-153.
6. Li, E., Chromatin modification and epigenetic reprogramming in mammalian development. *Nature Reviews Genetics* **2002,** *3* (9), 662-673.
7. Kwok, J., Role of epigenetics in Alzheimer's and Parkinson's disease. *Epigenomics* **2010,** *2* (5), 671–682.
8. Hedrich, C. M.; Tsokos, G. C., Epigenetic mechanisms in systemic lupus erythematosus and other autoimmune diseases. *Trends in Molecular Medicine* **2011,** *17* (12), 714-724.
9. Ordovás, J. M.; Smith, C. E., Epigenetics and cardiovascular disease. *Nature Reviews Cardiology* **2010,** *7* (9), 510-519.
10. Sarkies, P.; Sale, J. E., Cellular epigenetic stability and cancer. *Trends in Genetics* **2012,** *28* (3), 118-127.
11. Ogino, S.; Galon, J.; Fuchs, C. S.; Dranoff, G., Cancer immunology—analysis of host and tumor factors for personalized medicine. *Nature Reviews Clinical Oncology* **2011,** *8* (12), 711-719.
12. WATSON, J. D.; CRICK, F. H. C., Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953,** *171* (4356), 737-738.
13. Ghosh, A.; Bansal, M., A glossary of DNA structures from A to Z. *Acta Crystallographica. Section D* **2003,** *59* (Pt 4), 620-626.
14. Amanzadeh, E.; Mohabatkar, H.; Biria, D., Classification of DNA Minor and Major Grooves Binding Proteins According to the NLSs by Data Analysis Methods. *Applied Biochemistry and Biotechnology* **2014,** *174* (1), 437-451.
15. Oh, D. B.; Kim, Y. G.; Rich, A., Z-DNA-binding proteins can act as potent effectors of gene expression invivo. *Proceedings of the National Academy of Sciences* **2002,** *99* (26), 16666-16671.
16. Lodish, H.; Berk, A.; Zipursky, S. L.; Matsudaira, P.; Baltimore, D.; Darnell, J., *Molecular Cell Biology. 4th edition*. W. H. Freeman: 2000.
17. Reik, W., Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **2007,** *447*, 425–432.
18. Carell, T.; Germany, C. f. I. P. S. D. o. C. L.-M.-U. M. B.-M.; Kurz, M. Q.; Germany, C. f. I. P. S. D. o. C. L.-M.-U. M. B.-M.; Müller, M.; Germany, C. f. I. P. S. D. o. C. L.-M.-U. M. B.-M.; Rossa, M.; Germany, C. f. I. P. S. D. o. C. L.-M.-U. M. B.-M.; Spada, F.; Germany, C. f. I. P. S. D. o. C. L.-M.-U. M. B.-M., Non-canonical Bases in the Genome: The Regulatory Information Layer in DNA. *Angewandte Chemie International Edition* **2017,** *57* (16), 4296-4312.
19. Felsenfeld, G., A Brief History of Epigenetics. **2014,** (6), a018200.
20. Cheng, X., Structure and function of DNA methyltransferases. *Annual review of biophysics and biomolecular structure* **1995,** *24* (1), 293-318.
21. Meselson, M.; YuAn, R.; Heywood, J., Restriction and modification of DNA. *Annual review of biochemistry* **1972,** *41* (1), 447-466.
22. RA, M.; EJ, R., DNA Methylation in Eukaryotes. *Current opinion in genetics & development* **1995,** *5* (2), 234-242.
23. Blow, M. J.; Clark, T. A.; Daum, C. G.; Deutschbauer, A. M.; Fomenkov, A.; Fries, R.; Froula, J.; Kang, D. D.; Malmstrom, R. R.; Morgan, R. D.; Posfai, J.; Singh, K.; Visel, A.; Wetmore, K.; Zhao, Z.; Rubin, E. M.; Korlach, J.; Pennacchio, L. A.;

Roberts, R. J., The Epigenomic Landscape of Prokaryotes. *PLOS Genetics* **2016,** *12* (2), e1005854.
24. Li, E.; Zhang, Y., DNA Methylation in Mammals. **2014**.
25. Jabbari, K.; Bernardi, G., Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. *Gene* **2004,** *333*, 143-149.
26. Kohli, R. M.; Zhang, Y., TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* **2013,** *502* (7472), 472-479.
27. Tahiliani, M.; Koh, K. P.; Shen, Y.; Pastor, W. A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L. M.; Liu, D. R.; Aravind, L.; Rao, A., Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. **2009,** *324* (5929), 930-935.
28. Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y., Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. **2011,** *333* (6047), 1300-1303.
29. Maiti, A.; Drohat, A. C., Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine: POTENTIAL IMPLICATIONS FOR ACTIVE DEMETHYLATION OF CpG SITES*. *J Biol Chem* **2011,** *286* (41), 35334-35338.
30. Jang, H. S.; Shin, W. J.; Lee, J. E.; Do, J. T., CpG and Non-CpG Methylation in Epigenetic Gene Regulation and Brain Function. *Genes* **2017,** *8* (6), 148.
31. Jenuwein, T.; Allis, C. D., Translating the Histone Code. *Science* **2001,** *293* (5532), 1074-1080.
32. Campos, E. I.; Reinberg, D., Histones: Annotating Chromatin. *Annual Review of Genetics* **2009,** *43*, 559-599.
33. Arents, G.; Burlingame, R. W.; Wang, B. C.; Love, W. E.; Moudrianakis, E. N., The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. *Proc Natl Acad Sci* **1991,** *88* (22), 10148-52.
34. Jenuwein, T.; Allis, C. D., Translating the Histone Code. **2001,** *293* (5532), 1074-1080.
35. Strahl, B. D.; Allis, C. D., The language of covalent histone modifications. *Nature* **2000,** *403* (6765), 41-45.
36. Eberharter, A.; Becker, P. B., Histone acetylation: a switch between repressive and permissive chromatin: Second in review series on chromatin dynamics. *EMBO Rep* **2002,** *3* (3), 224-229.
37. Rossetto, D.; Avvakumov, N.; Côté, J., Histone phosphorylation: A chromatin modification involved in diverse nuclear events. *Epigenetics* **2012,** *7* (10), 1098-108.
38. Blanc, R. S.; Richard, S., Arginine Methylation: The Coming of Age: Molecular Cell. *Molecular Cell* **2017,** *65* (1), 8-24.
39. A.Gates, L.; E.Foulds, C.; W.O'Malley, B., Histone Marks in the 'Driver's Seat': Functional Roles in Steering the Transcription Cycle. *Trends in Biochemical Sciences* **2017,** *42* (12), 977-989.
40. Sims, R. J.; Reinberg, D., Histone H3 Lys 4 methylation: caught in a bind? *Genes & Development* **2006,** *20* (20), 2779-2786.
41. Guillemette, B.; Drogaris, P.; Lin, H. H.; Armstrong, H.; Hiragami-Hamada, K.; Imhof, A.; Bonneil, E.; Thibault, P.; Verreault, A.; Festenstein, R. J., H3 lysine 4 is acetylated at active gene promoters and is regulated by H3 lysine 4 methylation. *PLoS Genet* **2011,** *7* (3), e1001354.

42. Barski, A.; Cuddapah, S.; Cui, K.; Roh, T. Y.; Schones, D. E.; Wang, Z.; Wei, G.; Chepelev, I.; Zhao, K., High-resolution profiling of histone methylations in the human genome. *Cell* **2007,** *129* (4), 823-37.
43. Lee, T. I.; Young, R. A., TRANSCRIPTION OF EUKARYOTIC PROTEIN-CODING GENES. *Annual Reviews of Genetics* **2000,** *34*, 77-137.
44. Whitaker, J. W.; Chen, Z.; Wang, W., Predicting the human epigenome from DNA motifs. *Nature Methods* **2014,** *12* (3), 265-272.
45. Gardiner-Garden, M.; Frommer, M., CpG Islands in vertebrate genomes. *Journal of Molecular Biology* **1987,** *196* (2), 261-282.
46. Bird, A., DNA methylation patterns and epigenetic memory. *Genes & Development* **2002,** *16* (1), 6-21.
47. Arechederra, M.; Daian, F.; Yim, A.; Bazai, S. K.; Richelme, S.; Dono, R.; Saurin, A. J.; Habermann, B. H.; Maina, F., Hypermethylation of gene body CpG islands predicts high dosage of functional oncogenes in liver cancer. *Nature Communications* **2018,** *9* (1), 1-16.
48. Baylin, S. B.; Esteller, M.; Rountree, M. R.; Bachman, K. E.; Schuebel, K.; Herman, J. G., Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer. *Human Molecular Genetics* **2001,** *10* (7), 687-692.
49. Ng, H. H.; Bird, A., DNA methylation and chromatin modification. *Current Opinion in Genetic and Development* **1999,** *9* (2), 158-163.
50. Robertson, K. D.; Wolffe, A. P., DNA methylation in health and disease. *Nature Reviews Genetics* **2000,** *1* (1), 11-19.
51. Hellman, A.; Chess, A., Gene body-specific methylation on the active X chromosome. *Science* **2007,** *315* (5815), 1141-1143.
52. Jones, P. A., Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics* **2012,** *13* (7), 484-492.
53. de Koning, A. P. J.; Gu, W.; Castoe, T. A.; Batzer, M. A.; Pollock, D. D., Repetitive Elements May Comprise Over Two-Thirds of the Human Genome. *PLoS Genetics* **2011,** *7* (12), e102384.
54. Deniz, O.; Frost, J. M.; Branco, M. R., Regulation of transposable elements by DNA modifications. *Nature Reviews. Genetics.* **2019,** *20* (7), 417-431.
55. Rodic, N.; Sharma, R.; Zampella, J.; Dai, L.; Taylor, M. S.; Hruban, R. H.; Iacobuzio-Donahue, C. A.; Maitra, A.; Torbenson, M. S.; Goggins, M.; Shih Ie, M.; Duffield, A. S.; Montgomery, E. A.; Gabrielson, E.; Netto, G. J.; Lotan, T. L.; De Marzo, A. M.; Westra, W.; Binder, Z. A.; Orr, B. A.; Gallia, G. L.; Eberhart, C. G.; Boeke, J. D.; Harris, C. R.; Burns, K. H., Long interspersed element-1 protein expression is a hallmark of many human cancers. *The American Journal of Pathology* **2014,** *184* (5), 1280-1286.
56. Gemayel, R.; Cho, J.; Boeynaems, S.; Verstrepen, K. J., Beyond Junk-Variable Tandem Repeats as Facilitators of Rapid Evolution of Regulatory and Coding Sequences. *Genes* **2012,** *3* (3), 461-480.
57. Choi, S. H.; Worswick, S.; Byun, H. M.; Shear, T.; Soussa, J. C.; Wolff, E. M.; Douer, D.; Garcia-Manero, G.; Liang, G.; Yang, A. S., Changes in DNA methylation of tandem DNA repeats are different from interspersed repeats in cancer. *International Journal of Cancer* **2009,** *125* (3), 723-729.
58. Luger, K.; Mäder, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J., Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **1997,** *389* (6648), 251-260.

59. Thoma, F.; Koller, T., Influence of histone H1 on chromatin structure. *Cell* **1977,** *12* (1), 101-107.
60. Thoma, F.; Koller, T.; Klug, A., Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *The Journal of cell biology* **1979,** *83* (2 Pt 1), 403-427.
61. Berger, S. L., The complex language of chromatin regulation during transcription. *Nature* **2007,** *447* (7143), 407-412.
62. Fritz, A. J.; Sehgal, N.; Pliss, A.; Xu, J.; Berezney, R., Chromosome territories and the global regulation of the genome. *Genes Chromosomes Cancer* **2019,** *58* (7), 407-426.
63. Cremer, T.; Cremer, C., Rise, fall and resurrection of chromosome territories: a historical perspective. Part I. The rise of chromosome territories. *European Journal of Histochemistry* **2006,** *50* (3), 161-176.
64. Zorn, C.; Cremer, C.; Cremer, T.; Zimmer, J., Unscheduled DNA synthesis after partial UV irradiation of the cell nucleus. Distribution in interphase and metaphase. *Experimental Cell Research* **1979,** *124* (1), 111-119.
65. Cremer, T.; Peterson, S. P.; Cremer, C.; Berns, M. W., Laser microirradiation of Chinese hamster cells at wavelength 365 nm: effects of psoralen and caffeine. *Radiation Research* **1981,** *85* (3), 529-543.
66. Lanctôt, C.; Cheutin, T.; Cremer, M.; Cavalli, G.; Cremer, T., Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nature Reviews Genetics* **2007,** *8* (2), 104-115.
67. Stein, G. S.; Zaidi, S. K.; Stein, J. L.; Lian, J. B.; van Wijnen, A. J.; Montecino, M.; Young, D. W.; Javed, A.; Pratap, J.; Choi, J.-Y.; Ali, S. A.; Pande, S.; Hassan, M. Q., Genetic and epigenetic regulation in nuclear microenvironments for biological control in cancer. *Journal of Cellular Biochemistry* **2008,** *104* (6), 2016-2026.
68. Dixon, J. R.; Gorkin, D. U.; Ren, B., Chromatin Domains: The Unit of Chromosome Organization. *Molecular Cell* **2016,** *62* (5), 668-680.
69. Zink, D.; Cremer, T., Cell nucleus: chromosome dynamics in nuclei of living cells. *Current Biology* **1998,** *8* (9), R321-324.
70. Visser, A. E.; Aten, J. A., Chromosomes as well as chromosomal subdomains constitute distinct units in interphase nuclei. *Journal of Cell Science* **1999,** *112*, 3353-3360.
71. Kreth, G.; Finsterle, J.; von Hase, J.; Cremer, M.; Cremer, C., Radial arrangement of chromosome territories in human cell nuclei: a computer model approach based on gene density indicates a probabilistic global positioning code. *Biophysical Journal* **2004,** *86* (5), 2803-2812.
72. Sehgal, N.; Fritz, A. J.; Morris, K.; Torres, I.; Chen, Z.; Xu, J.; Berezney, R., Gene density and chromosome territory shape. *Chromosoma* **2014,** *123* (5), 499-513.
73. Caron, H.; van Schaik, B.; van der Mee, M.; Baas, F.; Riggins, G.; van Sluis, P.; Hermus, M. C.; van Asperen, R.; Boon, K.; Voute, P. A.; Heisterkamp, S.; van Kampen, A.; Versteeg, R., The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science* **2001,** *291* (5507), 1289-1292.
74. Grasser, F.; Neusser, M.; Fiegler, H.; Thormeyer, T.; Cremer, M.; Carter, N. P.; Cremer, T.; Müller, S., Replication-timing-correlated spatial chromatin arrangements in cancer and in primate interphase nuclei. *Journal of Cell Science* **2008,** *121*, 1876-1886.

75. Takizawa, T.; Meaburn, K. J.; Misteli, T., The meaning of gene positioning. *Cell* **2008,** *135* (1), 9-13.
76. Edelmann, P.; Bornfleth, H.; Zink, D.; Cremer, T.; Cremer, C., Morphology and dynamics of chromosome territories in living cells. *Biochimica et Biophysica Acta* **2001,** *1551* (1), M29-39.
77. Teller, K.; Illner, D.; Thamm, S.; Casas-Delucchi, C. S.; Versteeg, R.; Indemans, M.; Cremer, T.; Cremer, M., A top-down analysis of Xa- and Xi-territories reveals differences of higher order structure at >/= 20 Mb genomic length scales. *Nucleus* **2011,** *2* (5), 465-477.
78. Pliss, A.; Malyavantham, K.; Bhattacharya, S.; Zeitz, M.; Berezney, R., CHROMATIN DYNAMICS IS CORRELATED WITH REPLICATION TIMING. *Chromosoma* **2009,** *118* (4), 459-470.
79. Pliss, A.; Malyavantham, K. S.; Bhattacharya, S.; Berezney, R., Chromatin dynamics in living cells: identification of oscillatory motion. *Journal of Cellular Physiology* **2013,** *228* (3), 609-616.
80. Gasser, S. M., Visualizing chromatin dynamics in interphase nuclei. *Science* **2002,** *296* (5572), 1412-1416.
81. Oberdoerffer, P.; Sinclair, D. A., The role of nuclear architecture in genomic instability and ageing. *Nature Reviews. Molecular Cell Biology* **2007,** *8* (9), 692-702.
82. Nagano, T.; Lubling, Y.; Varnai, C.; Dudley, C.; Leung, W.; Baran, Y.; Mendelson Cohen, N.; Wingett, S.; Fraser, P.; Tanay, A., Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* **2017,** *547* (7661), 61-67.
83. Misteli, T., Beyond the Sequence: Cellular Organization of Genome Function. *Cell* **2007,** *128* (4), 787-800.
84. Stanek, D.; Fox, A. H., Nuclear bodies: news insights into structure and function. *Current Opinion in Cell Biology* **2017,** *46*, 94-101.
85. Stein, G. S.; Zaidi, S. K.; Braastad, C. D.; Montecino, M.; van Wijnen, A. J.; Choi, J. Y.; Stein, J. L.; Lian, J. B.; Javed, A., Functional architecture of the nucleus: organizing the regulatory machinery for gene expression, replication and repair. *Trends in Cell Biology* **2003,** *13* (11), 584-592.
86. Hnisz, D.; Shrinivas, K.; Young, R. A.; Chakraborty, A. K.; Sharp, P. A., A Phase Separation Model for Transcriptional Control. *Cell* **2017,** *169* (1), 13-23.
87. Erdel, F.; Rippe, K., Formation of Chromatin Subcompartments by Phase Separation. *Biophysical Journal* **2018,** *114* (10), 2262-2270.
88. Zink, D.; Fischer, A. H.; Nickerson, J. A., Nuclear structure in cancer cells. *Nature Reviews. Cancer.* **2004,** *4* (9), 677-687.
89. T, P., Half a century of "the nuclear matrix". *Molecular biology of the cell* **2000,** *11* (3), 799-805.
90. Arope, S.; Harraghy, N.; Pjanic, M.; Mermod, N., Molecular Characterization of a Human Matrix Attachment Region Epigenetic Regulator. *PLoS One* **2013,** *8* (11), e79262.
91. Lever, E.; Sheer, D., The role of nuclear organization in cancer. *The Journal of pathology* **2010,** *220* (2), 114-125.
92. Yang, J.; Corces, V. G., Chromatin Insulators: A Role in Nuclear Organization and Gene Expression. *Advances in Cancer Research* **2011,** *110*, 43-76.
93. Kadauke, S.; Blobel, G. A., Chromatin loops in gene regulation. *Biochimica et Biophysica Acta* **2009,** *1789* (1), 17-25.

94. Yu, M.; Ren, B., The Three-Dimensional Organization of Mammalian Genomes. *Annual Review of Cell and Developmental Biology* **2017,** *33*, 265-289.
95. Pombo, A.; Dillon, N., Three-dimensional genome architecture: players and mechanisms. *Nature Reviews Molecular Cell Biology* **2015,** *16* (4), 245-257.
96. Dixon, J. R.; Selvaraj, S.; Yue, F.; Kim, A.; Li, Y.; Shen, Y.; Hu, M.; Liu, J. S.; Ren, B., Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **2012,** *485* (7398), 376-380.
97. Mumbach, M. R.; Granja, J. M.; Flynn, R. A.; Roake, C. M.; Satpathy, A. T.; Rubin, A. J.; Qi, Y.; Jiang, Z.; Shams, S.; Louie, B. H.; Guo, J. K.; Gennert, D. G.; Corces, M. R.; Khavari, P. A.; Atianand, M. K.; Artandi, S. E.; Fitzgerald, K. A.; Greenleaf, W. J.; Chang, H. Y., HiChIRP reveals RNA-associated chromosome conformation. *Nature Methods* **2019,** *16* (6), 489-492.
98. Gruenbaum, Y.; Goldman, R. D.; Meyuhas, R.; Mills, E.; Margalit, A.; Fridkin, A.; Dayani, Y.; Prokocimer, M.; Enosh, A., The nuclear lamina and its functions in the nucleus. *International Review of Cytology* **2003,** *226*, 1-62.
99. Gonzalez-Sandoval, A.; Gasser, S. M., On TADs and LADs: Spatial Control Over Gene Expression. *Trends in Genetics* **2016,** *32* (8), 485-495.
100. Guelen, L.; Pagie, L.; Brasset, E.; Meuleman, W.; Faza, M. B.; Talhout, W.; Eussen, B. H.; de Klein, A.; Wessels, L.; de Laat, W.; van Steensel, B., Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **2008,** *453* (7197), 948-951.
101. van Koningsbruggen, S.; Gierliński, M.; Schofield, P.; Martin, D.; Barton, G. J.; Ariyurek, Y.; den Dunnen, J. T.; Lamond, A. I., High-Resolution Whole-Genome Sequencing Reveals That Specific Chromatin Domains from Most Human Chromosomes Associate with Nucleoli. *Colecular Biology of the Cell* **2010,** *21* (21), 3735-3748.
102. Iborra, F. J.; Pombo, A.; Jackson, D. A.; Cook, P. R., Active RNA polymerases are localized within discrete transcription “factories' in human nuclei. *Journal of Cell Science* **1996,** *109*, 1427-1436.
103. Jackson, D. A.; Hassan, A. B.; Errington, R. J.; Cook, P. R., Visualization of focal sites of transcription within human nuclei. *The EMBO Journal* **1993,** *12* (3), 1059-1065.
104. Weipoltshammer, K.; Schöfer, C., Morphology of nuclear transcription. *Histochemistry and Cell Biology* **2016,** *145* (4), 343-358.
105. Eskiw, C. H.; Rapp, A.; Carter, D. R. F.; Cook, P. R., RNA polymerase II activity is located on the surface of protein-rich transcription factories. *Journal of Cell Science* **2008,** *121*, 1999-2007.
106. Eskiw, C. H.; Fraser, P., Ultrastructural study of transcription factories in mouse erythroblasts. *Journal of Cell Science* **2011,** *124*, 3676-3683.
107. SK, P.; Y, X.; X, F.; WT, G., Pronounced Cohabitation of Active Immunoglobulin Genes From Three Different Chromosomes in Transcription Factories During Maximal Antibody Synthesis. *Genes & development* **2014,** *28* (11), 1159-1164.
108. II, C.; I, I.; SZ, C.; L, B.; A, S.; L, M.; C, D.-D.; B, H.; M, D.; X, D., Real-time dynamics of RNA polymerase II clustering in live human cells. *Science* **2013,** *341* (6146), 664-667.
109. Choy, J. S.; Wei, S.; Lee, J. Y.; Tan, S.; Chu, S.; Lee, T. H., DNA methylation increases nucleosome compaction and rigidity. *Journal of American Chemical Society* **2010,** *132* (6), 1782-1783.

110. Lee, J. Y.; Lee, T.-H., Effects of DNA Methylation on the Structure of Nucleosomes. *Journal of American Chemical Society* **2012,** *134* (1), 173-175.
111. Schoenherr, C. J.; Levorse, J. M.; Tilghman, S. M., CTCF maintains differential methylation at the Igf2/H19 locus. *Nature Genetics* **2003,** *33* (1), 66-69.
112. Flavahan, W. A.; Drier, Y.; Liau, B. B.; Gillespie, S. M.; Venteicher, A. S.; Stemmer-Rachamimov, A. O.; Suva, M. L.; Bernstein, B. E., Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **2016,** *529* (7584), 110-114.
113. Xie, W.; Ling, T.; Zhou, Y.; Feng, W.; Zhu, Q.; Stunnenberg, H. G.; Grummt, I.; Tao, W., The chromatin remodeling complex NuRD establishes the poised state of rRNA genes characterized by bivalent histone modifications and altered nucleosome positions. *Proceedings of the National Academy of Sciences* **2012,** *109* (21), 8161-8166.
114. J, E.; E, B.; R, S.; MF, F.; A, V.-G.; A, N.; L, L.-S.; S, R.; A, A.; H, O.; V, M.; A, J.; JC, S.; G, L.; I, G.; W, B.; M, E., Epigenetic Disruption of Ribosomal RNA Genes and Nucleolar Architecture in DNA Methyltransferase 1 (Dnmt1) Deficient Cells. *Nucleic acids research* **2007,** *35* (7), 2191-2198.
115. T, G.-K.; F, L.; V, S.; F, L.; T, M., Loss of Human Ribosomal Gene CpG Methylation Enhances Cryptic RNA Polymerase II Transcription and Disrupts Ribosomal RNA Processing. *Molecular cell* **2009,** *35* (4), 414-425.
116. Yanez-Cuna, J. O.; van Steensel, B., Genome-nuclear lamina interactions: from cell populations to single cells. *Current Opinion in Genetic and Development* **2017,** *43*, 67-72.
117. Dixon, J. R.; Jung, I.; Selvaraj, S.; Shen, Y.; Antosiewicz-Bourget, J. E.; Lee, A. Y.; Ye, Z.; Kim, A.; Rajagopal, N.; Xie, W.; Diao, Y.; Liang, J.; Zhao, H.; Lobanenkov, V. V.; Ecker, J. R.; Thomson, J. A.; Ren, B., Chromatin architecture reorganization during stem cell differentiation. *Nature* **2015,** *518* (7539), 331-336.
118. Maurano, M. T.; Humbert, R.; Rynes, E.; Thurman, R. E.; Haugen, E.; Wang, H.; Reynolds, A. P.; Sandstrom, R.; Qu, H.; Brody, J.; Shafer, A.; Neri, F.; Lee, K.; Kutyavin, T.; Stehling-Sun, S.; Johnson, A. K.; Canfield, T. K.; Giste, E.; Diegel, M.; Bates, D.; Hansen, R. S.; Neph, S.; Sabo, P. J.; Heimfeld, S.; Raubitschek, A.; Ziegler, S.; Cotsapas, C.; Sotoodehnia, N.; Glass, I.; Sunyaev, S. R.; Kaul, R.; Stamatoyannopoulos, J. A., Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science* **2012,** *337* (6099), 1190-1195.
119. Robertson, K. D., DNA methylation and human disease. *Nature Reviews Genetics* **2005,** *6* (8), 597-610.
120. Stirzaker, C.; Millar, D. S.; Paul, C. L.; Warnecke, P. M.; Harrison, J.; Vincent, P. C.; Frommer, M.; Clark, S. J., Extensive DNA methylation spanning the Rb promoter in retinoblastoma tumors. *Cancer Research* **1997,** *57* (11), 2229-2237.
121. Herman, J. G.; Merlo, A.; Mao, L.; Lapidus, R. G.; Issa, J. P.; Davidson, N. E.; Sidransky, D.; Baylin, S. B., Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers. *Cancer Research* **1995,** *55* (20), 4525-4530.
122. Robertson, K. D.; Jones, P. A., The Human ARF Cell Cycle Regulatory Gene Promoter Is a CpG Island Which Can Be Silenced by DNA Methylation and Down-Regulated by Wild-Type p53. *Molecular and Cellular Biology* **1998,** *18* (11), 6457-6473.
123. Moinova, H. R.; Chen, W. D.; Shen, L.; Smiraglia, D.; Olechnowicz, J.; Ravi, L.; Kasturi, L.; Myeroff, L.; Plass, C.; Parsons, R.; Minna, J.; Willson, J. K.; Green, S. B.;

Issa, J. P.; Markowitz, S. D., HLTF gene silencing in human colon cancer. *Proceedings of the National Academy of Sciences* **2002,** *99* (7), 4562-4567.
124. Issa, J. P.; Ottaviano, Y. L.; Celano, P.; Hamilton, S. R.; Davidson, N. E.; Baylin, S. B., Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon. *Nature Genetics* **1994,** *7* (4), 536-540.
125. Dobrovic, A.; Simpfendorfer, D., Methylation of the BRCA1 gene in sporadic breast cancer. *Cancer Research* **1997,** *57* (16), 3347-3350.
126. Katzenellenbogen, R. A.; Baylin, S. B.; Herman, J. G., Hypermethylation of the DAP-kinase CpG island is a common alteration in B-cell malignancies. *Blood* **1999,** *93* (12), 4347-4353.
127. Fontecave, M.; Atta, M.; Mulliez, E., S-adenosylmethionine: nothing goes to waste. *Trends in Biochemical Sciences* **2004,** *29* (5), 243-249.
128. Dalhoff, C.; Weinhold, E., S-Adenosyl-L-Methionine and Related Compounds. *Modified Nucleosides: in Biochemistry, Biotechnology and Medicine* **2008**, 223-247.
129. Struck, A.-W.; Thompson, M. L.; Wong, L. S.; Micklefield, J., S-Adenosyl-Methionine-Dependent Methyltransferases: Highly Versatile Enzymes in Biocatalysis, Biosynthesis and Other Biotechnological Applications. *ChemBioChem* **2012,** *13* (18), 2642-2655.
130. Frey, P. A.; Magnusson, O. T., S-Adenosylmethionine: A Wolf in Sheep's Clothing, or a Rich Man's Adenosylcobalamin? *Chemical Reviews* **2003,** *103* (6), 2129-2148.
131. Loenen, W. A. M., S-Adenosylmethionine: jack of all trades and master of everything? *Biochemical Society Transactions* **2006,** *34* (2), 330.
132. Martin, J. L.; McMillan, F. M., SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Current Opinion in Structural Biology* **2002,** *12* (6), 783-793.
133. Craddock, N.; Owen, M. J.; O'Donovan, M. C., The catechol-O-methyl transferase (COMT) gene as a candidate for psychiatric phenotypes: evidence and lessons. *Mol Psychiatry* **2006,** *11* (5), 446-458.
134. Locasale, J. W., Serine, glycine and one-carbon units: cancer metabolism in full circle. *Nat Rev Cancer* **2013,** *13* (8), 572-583.
135. Agúndez, J. A. G.; Luengo, A.; Herráez, O.; Martínez, C.; Alonso-Navarro, H.; Jiménez-Jiménez, F. J.; García-Martín, E., Nonsynonymous Polymorphisms of Histamine-Metabolising Enzymes in Patients with Parkinson's Disease. *NeuroMolecular Medicine* **2008,** *10* (1), 10-16.
136. Sharma, S.; Mann, D.; Singh, T. P.; Ghosh, B., Lack of association of histamine-N-methyltransferase (HNMT) polymorphisms with asthma in the Indian population. *Journal of Human Genetics* **2005,** *50* (12), 611-617.
137. Corrales, R. M.; Molle, V.; Leiba, J.; Mourey, L.; de Chastellier, C.; Kremer, L., Phosphorylation of mycobacterial PcaA inhibits mycolic acid cyclopropanation: consequences for intracellular survival and for phagosome maturation block. *The Journal of biological chemistry* **2012,** *287* (31), 26187-26199.
138. Boriack-Sjodin, P. A.; Swinger, K. K., Protein Methyltransferases: A Distinct, Diverse, and Dynamic Family of Enzymes. *Biochemistry* **2016,** *55* (11), 1557-1569.
139. Egloff, M. P.; Benarroch, D.; Selisko, B.; Romette, J. L.; Canard, B., An RNA cap (nucleoside-2′-O-)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization. *The EMBO Journal* **2002,** *21* (11), 2757.

140. Seidel-Rogol, B. L.; McCulloch, V.; Shadel, G. S., Human mitochondrial transcription factor B1 methylates ribosomal RNA at a conserved stem-loop. *Nature genetics* **2003,** *33* (1).
141. Song, J.; Rechkoblit, O.; Bestor, T. H.; Patel, D. J., Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science* **2011,** *331* (6020), 1036-1040.
142. Okano, M.; Xie, S.; Li, E., Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nature Genetics* **1998,** *19* (3), 219-220.
143. Palmer, J. L.; Abeles, R. H., The mechanism of action of S-adenosylhomocysteinase. *Journal of Biological Chemistry* **1979,** *254* (4), 1217-1226.
144. M.L., L., STRUCTURE-BASED PERSPECTIVES ON B12-DEPENDENT ENZYMES. *Annual Review of Biochemistry* **1997,** *66* (1), 269-313.
145. Pignot, M.; Siethoff, C.; Linscheid, M.; Weinhold, E., Coupling of a nucleoside with DNA by a methyltransferase. *Angewandte Chemie International Edition* **1998,** *37* (20), 2888-2891.
146. Pljevaljčić, G.; Schmidt, F.; Weinhold, E., Sequence-specific Methyltransferase-Induced Labeling of DNA (SMILing DNA). *ChemBioChem* **2004,** *5* (3), 265-269.
147. Pljevaljcic, G.; Pignot, M.; Weinhold, E., Design of a New Fluorescent Cofactor for DNA Methyltransferases and Sequence-Specific Labeling of DNA. *Journal of the American Chemical Society* **2003,** *125* (12), 3486-3492.
148. Comstock, L. R.; Rajski, S. R., Conversion of DNA methyltransferases into azidonucleosidyl transferases via synthetic cofactors. *Nucleic Acids Research* **2005,** *33* (5), 1644-1652.
149. Weller, R. L.; Rajski, S. R., DNA Methyltransferase-Moderated Click Chemistry. *Organic Letters* **2005,** *7* (11), 2141-2144.
150. Zhang, J.; Zheng, Y. G., SAM/SAH Analogs as Versatile Tools for SAM-Dependent Methyltransferases. *ACS Chemical Biology* **2016,** *11* (3), 583-597.
151. Dalhoff, C.; Lukinavicius, G.; Klimasauskas, S.; Weinhold, E., Synthesis of S-adenosyl-L-methionine analogs and their use for sequence-specific transalkylation of DNA by methyltransferases. *Nat. Protocols* **2006,** *1* (4), 1879-1886.
152. Lukinavičius, G.; Tomkuvienė, M.; Masevičius, V.; Klimašauskas, S., Enhanced Chemical Stability of AdoMet Analogues for Improved Methyltransferase-Directed Labeling of DNA. *ACS Chemical Biology* **2013,** *8* (6), 1134-1139.
153. Singh, S.; Zhang, J.; Huber, T. D.; Sunkara, M.; Hurley, K.; Goff, R. D.; Wang, G.; Zhang, W.; Liu, C.; Rohr, J.; Van Lanen, S. G.; Morris, A. J.; Thorson, J. S., Facile Chemoenzymatic Strategies for the Synthesis and Utilization of S-Adenosyl-L-Methionine Analogues. *Angewandte Chemie International Edition* **2014,** *53* (15), 3965-3969.
154. Lipson, J. M.; Thomsen, M.; Moore, B. S.; Clausen, R. P.; La Clair, J. J.; Burkart, M. D., A Tandem Chemoenzymatic Methylation by S-Adenosyl-L-methionine. *ChemBioChem* **2013,** *14* (8), 950-953.
155. Lukinavičius, G.; Lapienė, V.; Staševskij, Z.; Dalhoff, C.; Weinhold, E.; Klimašauskas, S., Targeted Labeling of DNA by Methyltransferase-Directed Transfer of Activated Groups (mTAG). *Journal of the American Chemical Society* **2007,** *129* (10), 2758-2759.
156. Lee, B. W. K.; Sun, H. G.; Zang, T.; Kim, B. J.; Alfaro, J. F.; Zhou, Z. S., Enzyme-Catalyzed Transfer of a Ketone Group from an S-Adenosylmethionine Analogue: A

Tool for the Functional Analysis of Methyltransferases. *Journal of the American Chemical Society* **2010,** *132* (11), 3642-3643.
157. Deen, J.; Vranken, C.; Leen, V.; Neely, R. K.; Janssen, K. P. F.; Hofkens, J., Methyltransferase directed labeling of biomolecules and its applications. *Angewandte Chemie International Edition* **2016**, n/a-n/a.
158. Lin, Q.; Jiang, F.; Schultz, P. G.; Gray, N. S., Design of Allele-Specific Protein Methyltransferase Inhibitors. *Journal of the American Chemical Society* **2001,** *123* (47), 11608-11613.
159. Loenen, W. A. M.; Dryden, D. T. F.; Raleigh, E. A.; Wilson, G. G.; Murray, N. E., Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic Acids Research* **2014,** *42* (1), 3-19.
160. Kriukiene, E.; Labrie, V.; Khare, T.; Urbanaviciute, G.; Lapinaite, A.; Koncevicius, K.; Li, D.; Wang, T.; Pai, S.; Ptak, C.; Gordevicius, J.; Wang, S. C.; Petronis, A.; Klimasauskas, S., DNA unmethylome profiling by covalent capture of CpG sites. *Nat Commun* **2013,** *4*, 2190.
161. Campbell, R. M.; Tummino, P. J., Cancer epigenetics drug discovery and development: the challenge of hitting the mark. *The Journal of Clinical Investigation* **2014,** *124* (1), 64-69.
162. Richon, V. M.; Johnston, D.; Sneeringer, C. J.; Jin, L.; Majer, C. R.; Elliston, K.; Jerva, L. F.; Scott, M. P.; Copeland, R. A., Chemogenetic Analysis of Human Protein Methyltransferases. *Chemical Biology & Drug Design* **2011,** *78* (2), 199-210.
163. Devkota, K.; Lohse, B.; Liu, Q.; Wang, M. W.; Stærk, D.; Berthelsen, J.; Clausen, R. P., Analogues of the Natural Product Sinefungin as Inhibitors of EHMT1 and EHMT2. *ACS Medicinal Chemistry Letters* **2014,** *5* (4), 293-297.
164. Zheng, W.; Ibáñez, G.; Wu, H.; Blum, G.; Zeng, H.; Dong, A.; Li, F.; Hajian, T.; Allali-Hassani, A.; Amaya, M. F.; Siarheyeva, A.; Yu, W.; Brown, P. J.; Schapira, M.; Vedadi, M.; Min, J.; Luo, M., Sinefungin Derivatives as Inhibitors and Structure Probes of Protein Lysine Methyltransferase SETD2. *Journal of the American Chemical Society* **2012,** *134* (43), 18004-18014.
165. D, S.; MS, E.; F, L.; S, K.; MM, S.; PJ, B.; D, B.-L.; CH, A.; M, V.; M, S., Discovery of a Dual PRMT5-PRMT7 Inhibitor. *ACS Medicinal Chemistry Letters* **2015,** *6* (4), 408-412.
166. C, S.; F, L., Modes of Action of the DNA Methyltransferase Inhibitors Azacytidine and Decitabine. *International journal of cancer* **2008,** *123* (1), 8-13.
167. OM, S.; L, I.; DB, L.; L, Z.; N, B.; S, C.; F, R.; A, V.; N, E.; AJ, P.; J, R.; N, B.; AR, M.; D, D.; JM, B.; A, W., SAR Around (l)-S-adenosyl-l-homocysteine, an Inhibitor of Human DNA Methyltransferase (DNMT) Enzymes. *Bioorganic & medicinal chemistry letters* **2009,** *19* (10), 2747-2751.
168. Lim, S. P.; Sonntag, L. S.; Noble, C.; Nilar, S. H.; Ng, R. H.; Zou, G.; Monaghan, P.; Chung, K. Y.; Dong, H.; Liu, B.; Bodenreider, C.; Lee, G.; Ding, M.; Chan, W. L.; Wang, G.; Jian, Y. L.; Chao, A. T.; Lescar, J.; Yin, Z.; Vedananda, T. R.; Keller, T. H.; Shi, P. Y., Small Molecule Inhibitors That Selectively Block Dengue Virus Methyltransferase*. *Journal of Biological Chemistry* **2011,** *286* (8), 6233-6240.
169. W, A.; C, E.; B, C.; B, M.; F, D.; JJ, V.; JM, C.; C, M.; G, Q.; ML, J.; B, C.; JC, G.; E, D., Toward the Identification of Viral Cap-Methyltransferase Inhibitors by Fluorescence Screening Assay. *Antiviral research* **2017,** *144*, 330-339.
170. Crewe, A. V.; Wall, J.; Langmore, J., Visibility of Single Atoms. *Science* **1970,** *168* (3937), 1338-1340.

171. Lichtman, J. W.; Conchello, J.-A., Fluorescence microscopy. *Nature Methods* **2005,** *2* (12), 910-919.
172. Monici, M., Cell and tissue autofluorescence research and diagnostic applications. *Biotechnology Annual Review* **2005,** *11*, 227-256.
173. Marchesini, R.; Brambilla, M.; Pignoli, E.; Bottiroli, G.; Croce, A. C.; Dal Fante, M.; Spinelli, P.; di Palma, S., Light-induced fluorescence spectroscopy of adenomas, adenocarcinomas and non-neoplastic mucosa in human colon. I. In vitro measurements. *Journal of Photochemistry and Photobiology* **1992,** *14* (3), 219-230.
174. Koenig, F.; McGovern, F. J.; Althausen, A. F.; Deutsch, T. F.; Schomacker, K. T., Laser induced autofluorescence diagnosis of bladder cancer. *The Journal of Urology* **1996,** *156* (5), 1597-1601.
175. Zheng, Q.; Lavis, L. D., Development of photostable fluorophores for molecular imaging. *Current Opinion in Chemical Biology* **2017,** *39*, 32-38.
176. Raulf, A.; Spahn, C. K.; Zessin, P. J. M.; Finan, K.; Bernhardt, S.; Heckel, A.; Heilemann, M., Click chemistry facilitates direct labelling and super-resolution imaging of nucleic acids and proteins. *RSC Advances* **2014,** (57), 30462-30466.
177. Toseland, C. P., Fluorescent labeling and modification of proteins. *Journal of Chemical Biology* **2013,** *6* (3), 85-95.
178. Tarnowski, B. I.; Spinale, F. G.; Nicholson, J. H., DAPI as a useful stain for nuclear quantitation. *Biotech Histochem* **1991,** *66* (6), 297-302.
179. Jr., R. C. S.; Grotyohann, L. W., Measurement of Mitochondrial Membrane Potential Using Fluorescent Rhodamine Derivatives. *Biophysical Journal* **1999,** *76* (1), 469-477.
180. Yapici, N. B.; Bi, Y.; Li, P.; Chen, X.; Yan, X.; Mandalapu, S. R.; Faucett, M.; Jockusch, S.; Ju, J.; Gibson, K. M.; Pavan, W. J.; Bi, L., Highly Stable and Sensitive Fluorescent Probes (LysoProbes) for Lysosomal Labeling and Tracking. *Scientific Reports* **2015,** *5* (1), 1-8.
181. Rudolf, R.; Mongillo, M.; Rizzuto, R.; Pozzan, T., Looking forward to seeing calcium. *Nature Reviews. Molecular Cell Biology* **2003,** *4* (7), 579-586.
182. Bassnett, S.; Reinisch, L.; Beebe, D. C., Intracellular pH measurement using single excitation-dual emission fluorescence ratios. *The American Journal of Physiology* **1990,** *258* (1), C171-C178.
183. Peterka, D. S.; Takahashi, H.; Yuste, R., Imaging voltage in neurons. *Neuron* **2011,** *69* (1), 9-21.
184. Shimomura, O.; Johnson, F. H.; Saiga, Y., Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. *Journal of Cellular and Comparative Physiology* **1962,** *59*, 223-239.
185. Shaner, N. C.; Campbell, R. E.; Steinbach, P. A.; Giepmans, B. N. G.; Palmer, A. E.; Tsien, R. Y., Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nature Biotechnology* **2004,** *22* (12), 1567-1572.
186. Cubitt, A. B.; Heim, R.; Adams, S. R.; Boyd, A. E.; Gross, L. A.; Tsien, R. Y., Understanding, improving and using green fluorescent proteins. *Trends in Biochemical Sciences* **1995,** *20* (11), 448-455.
187. Kremers, G.-J.; Gilbert, S. G.; Cranfill, P. J.; Davidson, M. W.; Piston, D. W., Fluorescent proteins at a glance. *Journal of Cell Science* **2011,** (124), 157-160.
188. Sanderson, M. J.; Smith, I.; Parker, I.; Bootman, M. D., Fluorescence microscopy. *Cold Spring Harbor Protocols* **2014,** *2014* (10), pdb.top071795.

189. Huang, B.; Bates, M.; Zhuang, X., Super-resolution fluorescence microscopy. *Annual Review of Biochemistry* **2009,** *78*, 993-1016.
190. Swedlow, J. R., Quantitative fluorescence microscopy and image deconvolution. *Methods in Cell Biology* **2013,** *114*, 407-426.
191. Kino, T. R. C. a. G. S., *Confocal Scanning Optical Microscopy and Related Imaging Systems*. Academic Press: 1996.
192. Sage, D.; Donati, L.; Soulez, F.; Fortun, D.; Schmit, G.; Seitz, A.; Guiet, R.; Vonesch, C.; Unser, M., DeconvolutionLab2: An open-source software for deconvolution microscopy. *Methods* **2017,** *115*, 28-41.
193. McNally, J. G.; Karpova, T.; Cooper, J.; Conchello, J. A., Three-dimensional imaging by deconvolution microscopy. *Methods* **1999,** *19* (3), 373-385.
194. Huisken, J.; Swoger, J.; Bene, F. D.; Wittbrodt, J.; Stelzer, E. H. K., Optical Sectioning Deep Inside Live Embryos by Selective Plane Illumination Microscopy. *Science* **2004,** *305* (5686), 1007-1009.
195. Keller, P. J.; Schmidt, A. D.; Wittbrodt, J.; Stelzer, E. H. K., Reconstruction of Zebrafish Early Embryonic Development by Scanned Light Sheet Microscopy. *Science* **2008,** *322* (5904), 1065-1069.
196. Verveer, P. J.; Swoger, J.; Pampaloni, F.; Greger, K.; Marcello, M.; Stelzer, E. H., High-resolution three-dimensional imaging of large specimens with light sheet-based microscopy. *Nature Methods* **2007,** *4* (4), 311-313.
197. Wu, Y.; Ghitani, A.; Christensen, R.; Santella, A.; Du, Z.; Rondeau, G.; Bao, Z.; Colón-Ramos, D.; Shroff, H., Inverted selective plane illumination microscopy (iSPIM) enables coupled cell identity lineaging and neurodevelopmental imaging in Caenorhabditis elegans. *Proceedings of the National Academy of Sciences* **2011,** *108* (43), 17708-17713.
198. Schmid, B.; Shah, G.; Scherf, N.; Weber, M.; Thierbach, K.; Campos, C. P.; Roeder, I.; Aanstad, P.; Huisken, J., High-speed panoramic light-sheet microscopy reveals global endodermal cell dynamics. *Nature Communications* **2013,** *4* (1), 1-10.
199. Galland, R.; Grenci, G.; Aravind, A.; Viasnoff, V.; Studer, V.; Sibarita, J.-B., 3D high- and super-resolution imaging using single-objective SPIM. *Nature Methods* **2015,** *12* (7), 641-644.
200. Zanacchi, F. C.; Lavagnino, Z.; Donnorso, M. P.; Bue, A. D.; Furia, L.; Faretta, M.; Diaspro, A., Live-cell 3D super-resolution imaging in thick biological samples. *Nature Methods* **2011,** *8* (12), 1047-1049.
201. Cella Zanacchi, F.; Lavagnino, Z.; Faretta, M.; Furia, L.; Diaspro, A., Light-Sheet Confined Super-Resolution Using Two-Photon Photoactivation. *PLoS One* **2013,** *8* (7), e67667.
202. James, P., *Handbook of Biological Confocal Microscopy | James Pawley | Springer*. New York: Springer: 2006.
203. Bailey, B.; Farkas, D. L.; Taylor, D. L.; Lanni, F., Enhancement of axial resolution in fluorescence microscopy by standing-wave excitation. *Nature* **1993,** *366* (6450), 44-48.
204. Gustafsson, M. G., Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *Journal of microscopy* **2000,** *198* (2), 82-87.
205. Gustafsson, M. G.; Shao, L.; Carlton, P. M.; Wang, C. J.; Golubovskaya, I. N.; Cande, W. Z.; Agard, D. A.; Sedat, J. W., Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination. *Biophysical Journal* **2008,** *94* (12), 4957-4970.

206. Schermelleh, L.; Carlton, P. M.; Haase, S.; Shao, L.; Winoto, L.; Kner, P.; Burke, B.; Cardoso, M. C.; Agard, D. A.; Gustafsson, M. G.; Leonhardt, H.; Sedat, J. W., Subdiffraction multicolor imaging of the nuclear periphery with 3D structured illumination microscopy. *Science* **2008,** *320* (5881), 1332-1336.
207. Ha, T.; Tinnefeld, P., Photophysics of Fluorescence Probes for Single Molecule Biophysics and Super-Resolution Imaging. *Annual Review of Physical Chemistry* **2012,** *63*, 595-617.
208. Rust, M. J.; Bates, M.; Zhuang, X., Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods* **2006,** *3* (10), 793-795.
209. Betzig, E.; Patterson, G. H.; Sougrat, R.; Lindwasser, O. W.; Olenych, S.; Bonifacino, J. S.; Davidson, M. W.; Lippincott-Schwartz, J.; Hess, H. F., Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **2006,** *313* (5793), 1642-1645.
210. Hess, S. T.; Girirajan, T. P.; Mason, M. D., Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophysical Journal* **2006,** *91* (11), 4258-4272.
211. Shim, S. H.; Xia, C.; Zhong, G.; Babcock, H. P.; Vaughan, J. C.; Huang, B.; Wang, X.; Xu, C.; Bi, G. Q.; Zhuang, X., Super-resolution fluorescence imaging of organelles in live cells with photoswitchable membrane probes. *Proceedings of the National Academy of Sciences* **2012,** *109* (35), 13978-13983.
212. Wu, Y.; Wu, X.; Lu, R.; Zhang, J.; Toro, L.; Stefani, E., Resonant Scanning with Large Field of View Reduces Photobleaching and Enhances Fluorescence Yield in STED Microscopy. *Scientific Reports* **2015,** *5* (1), 1-12.
213. Schermelleh, L.; Heintzmann, R.; Leonhardt, H., A guide to super-resolution fluorescence microscopy. *The Journal of Cell Biology* **2010,** *190* (2), 165-175.
214. Bates, M.; Dempsey, G. T.; Chen, K. H.; Zhuang, X., Multicolor super-resolution fluorescence imaging via multi-parameter fluorophore detection. *Chemphyschem* **2012,** *13* (1), 99-107.
215. Dalhoff, C.; Hüben, M.; Lenz, T.; Poot, P.; Nordhoff, E.; Köster, H.; Weinhold, E., Synthesis of S-Adenosyl-L-homocysteine Capture Compounds for Selective Photoinduced Isolation of Methyltransferases. *ChemBioChem* **2010,** *11* (2), 256-265.
216. Wang, R.; Zheng, W.; Yu, H.; Deng, H.; Luo, M., Labeling Substrates of Protein Arginine Methyltransferase with Engineered Enzymes and Matched S-Adenosyl-l-methionine Analogues. *Journal of the American Chemical Society* **2011,** *133* (20), 7648-7651.
217. Wang, R.; Islam, K.; Liu, Y.; Zheng, W.; Tang, H.; Lailler, N.; Blum, G.; Deng, H.; Luo, M., Profiling Genome-Wide Chromatin Methylation with Engineered Posttranslation Apparatus within Living Cells. *Journal of American Chemical Society* **2013,** *135* (3), 1048-1056.
218. Hymbaugh Bergman, S. J.; Comstock, L. R., N-mustard analogs of S-adenosyl-l-methionine as biochemical probes of protein arginine methylation. *Bioorganic & Medicinal Chemistry* **2015,** *23* (15), 5050-5055.
219. Andrew Streitwieser, J., Solvolytic Displacement Reactions At Saturated Carbon Atoms. *Chemical Reviews* **1956,** *56* (4), 571-752.
220. Sun, H.; Yeo, W. L.; Lim, Y. H.; Chew, X.; Smith, D. J.; Xue, B.; Chan, K. P.; Robinson, R. C.; Robins, E. G.; Zhao, H.; Ang, E. L., Directed Evolution of a Fluorinase for Improved Fluorination Efficiency with a Non-native Substrate. *Angewandte Chemie International Edition* **2016,** *55* (46), 14277-14280.

221. Ji, X.; Li, Y.; Xie, L.; Lu, H.; Ding, W.; Zhang, Q., Expanding Radical SAM Chemistry by Using Radical Addition Reactions and SAM Analogues. *Angewandte Chemie International Edition* **2016,** *55* (39), 11845-11848.
222. Dalhoff, C.; Lukinavičius, G.; Klimašauskas, S.; Weinhold, E., Direct transfer of extended groups from synthetic cofactors by DNA methyltransferases. *Nature Chemical Biology* **2005,** *2* (1), 31-32.
223. Clark, S. J.; Lee, H. J.; Smallwood, S. A.; Kelsey, G.; Reik, W., Single-cell epigenomics: powerful new methods for understanding gene regulation and cell identity. *Genome Biology* **2016,** *17* (1), 1-10.
224. Booth, M. J.; Raiber, E.-A.; Balasubramanian, S., Chemical Methods for Decoding Cytosine Modifications in DNA. *Chemical Reviews* **2015,** *115* (6), 2240-2254.
225. Consortium, T. E. P., An Integrated Encyclopedia of DNA Elements in the Human Genome. *Nature* **2012,** *489* (7414), 57-74.
226. Bheda, P.; Schneider, R., Epigenetics reloaded: the single-cell revolution. *Trends in Cell Biology* **2014,** *24* (11), 712-723.
227. Macaulay, I. C.; Voet, T., Single Cell Genomics: Advances and Future Perspectives. *PLoS Genetics* **2014,** *10* (1), e1004126.
228. Cokus, S. J.; Feng, S.; Zhang, X.; Chen, Z.; Merriman, B.; Haudenschild, C. D.; Pradhan, S.; Nelson, S. F.; Pellegrini, M.; Jacobsen, S. E., Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* **2008,** *452* (7184), 215-219.
229. M, W.; JJ, D.; D, W.; EJ, O.; M, H.; WL, L.; D, S., Chromosome-wide and Promoter-Specific Analyses Identify Sites of Differential DNA Methylation in Normal and Transformed Human Cells. *Nature genetics* **2005,** *37* (8), 853-862.
230. Park, P. J., ChIP–seq: advantages and challenges of a maturing technology. *Nature Reviews Genetics* **2009,** *10* (10), 669-680.
231. Robinson, M. D.; Stirzaker, C.; Statham, A. L.; Coolen, M. W.; Song, J. Z.; Nair, S. S.; Strbenac, D.; Speed, T. P.; Clark, S. J., Evaluation of affinity-based genome-wide DNA methylation data: Effects of CpG density, amplification bias, and copy number variation. *Genome Research* **2010,** *20* (12), 1719-1729.
232. A, A.; HG, M.; None, A.; X, X.; JO, K.; EH, T.; B, S.; AP, M.; NC, C.; X, Z.; J, S., Rapid, Low-Input, Low-Bias Construction of Shotgun Fragment Libraries by High-Density in Vitro Transposition. *Genome biology* **2010,** *11* (12), R119.
233. W, J.; Q, T.; M, W.; K, C.; Y, Z.; G, R.; B, N.; J, S.; TM, P.; R, C.; D, L.; K, Z., Genome-wide Detection of DNase I Hypersensitive Sites in Single Cells and FFPE Tissue Samples. *Nature* **2015,** *528* (7580), 142-146.
234. Dekker, J.; Rippe, K.; Dekker, M.; Kleckner, N., Capturing Chromosome Conformation. *Science* **2002,** *295* (5558), 1306-1311.
235. Lieberman-Aiden, E.; van Berkum, N. L.; Williams, L.; Imakaev, M.; Ragoczy, T.; Telling, A.; Amit, I.; Lajoie, B. R.; Sabo, P. J.; Dorschner, M. O.; Sandstrom, R.; Bernstein, B.; Bender, M. A.; Groudine, M.; Gnirke, A.; Stamatoyannopoulos, J.; Mirny, L. A.; Lander, E. S.; Dekker, J., Comprehensive mapping of long range interactions reveals folding principles of the human genome. *Science* **2009,** *326* (5950), 289-293.
236. Nagano, T.; Lubling, Y.; Stevens, T. J.; Schoenfelder, S.; Yaffe, E.; Dean, W.; Laue, E. D.; Tanay, A.; Fraser, P., Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **2013,** *502* (7469), 59-64.

237. Stevens, T. J.; Lando, D.; Basu, S.; Atkinson, L. P.; Cao, Y.; Lee, S. F.; Leeb, M.; Wohlfahrt, K. J.; Boucher, W.; O'Shaughnessy-Kirwan, A.; Cramard, J.; Faure, A. J.; Ralser, M.; Blanco, E.; Morey, L.; Sansó, M.; Palayret, M. G. S.; Lehner, B.; Croce, L. D.; Wutz, A.; Hendrich, B.; Klenerman, D.; Laue, E. D., 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **2017,** *544* (7648), 59-64.
238. Li, G.; Liu, Y.; Zhang, Y.; Kubo, N.; Yu, M.; Fang, R.; Kellis, M.; Ren, B., Joint profiling of DNA methylation and chromatin architecture in single cells. *Nature Methods* **2019,** *16* (10), 991-993.
239. Pal, K.; Forcato, M.; Ferrari, F., Hi-C analysis: from data generation to integration. *Biophysical Reviews* **2018,** *11* (1), 67-78.
240. AL, O.; DE, O., Spheroid Chromatin Units (V Bodies). *Science* **1974,** *183* (4122), 330-332.
241. J, C.; S, A.; JC, C.; E, H., Combined Immunofluorescence, RNA Fluorescent in Situ Hybridization, and DNA Fluorescent in Situ Hybridization to Study Chromatin Changes, Transcriptional Activity, Nuclear Organization, and X-chromosome Inactivation. *Methods in Molecular Biology* **2008,** *463*, 297-308.
242. Y, L.; Y, M.; K, S.; H, N.; T, K.; H, O.; A, O.; H, S., Sequence-specific Microscopic Visualization of DNA Methylation Status at Satellite Repeats in Individual Cell Nuclei and Chromosomes. *Nucleic acids research* **2013,** *41* (19), e186.
243. GJ, N.; TW, P.; SA, B.; SB, B.; JG, H., In Situ Detection of the Hypermethylation-Induced Inactivation of the p16 Gene as an Early Event in Oncogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **1999,** *96* (22), 12754-12759.
244. Bergqvist, C.; Department of Biochemistry and Biophysics, S. U., Svante Arrhenius väg 16B, SE-106 91 Stockholm, Sweden; Niss, F.; Department of Biochemistry and Biophysics, S. U., Svante Arrhenius väg 16B, SE-106 91 Stockholm, Sweden; Figueroa, R. A.; Department of Biochemistry and Biophysics, S. U., Svante Arrhenius väg 16B, SE-106 91 Stockholm, Sweden; Beckman, M.; Department of Biochemistry and Biophysics, S. U., Svante Arrhenius väg 16B, SE-106 91 Stockholm, Sweden; Institute of Environmental Medicine, K. I. S.-S.; Maksel, D.; Monash Molecular Crystallisation Facility (MMCF), M. U., VIC 3800, Australia; Jafferali, M. H.; Department of Biochemistry and Biophysics, S. U., Svante Arrhenius väg 16B, SE-106 91 Stockholm, Sweden; Kulyté, A.; Lipid laboratory, D. o. M., Karolinska Institutet, SE-141 57 Huddinge, Sweden; Ström, A.-L.; Department of Biochemistry and Biophysics, S. U., Svante Arrhenius väg 16B, SE-106 91 Stockholm, Sweden; Hallberg, E.; Department of Biochemistry and Biophysics, S. U., Svante Arrhenius väg 16B, SE-106 91 Stockholm, Sweden, Monitoring of chromatin organization in live cells by FRIC. Effects of the inner nuclear membrane protein Samp1. *Nucleic Acids Research* **2019,** *47* (9), e49.
245. Y, S.; T, K.; R, A.; H, A.; C, O.; N, H.; K, Y.; J, U.; T, N.; T, H.; Y, H.; A, K.; H, K.; H, K., A Genetically Encoded Probe for Live-Cell Imaging of H4K20 Monomethylation. *Journal of molecular biology* **2016,** *428* (20), 3885-3902.
246. Stasevich, T. J.; Hayashi-Takanaka, Y.; Sato, Y.; Maehara, K.; Ohkawa, Y.; Sakata-Sogawa, K.; Tokunaga, M.; Nagase, T.; Nozaki, N.; McNally, J. G.; Kimura, H., Regulation of RNA polymerase II activation by histone acetylation in single living cells. *Nature* **2014,** *516* (7530), 272-275.

247. Lungu, C.; Pinter, S.; Broche, J.; Rathert, P.; Jeltsch, A., Modular fluorescence complementation sensors for live cell detection of epigenetic signals at endogenous genomic sites. *Nature Communications* **2017,** *8* (1), 1-13.
248. E, T.; NM, L.; RH, S.; B, W., Imaging mRNA In Vivo, from Birth to Death. *Annual review of biophysics* **2018,** *47*, 85-106.
249. H, S.; S, D.; RH, S.; M, V., Imaging of DNA and RNA in Living Eukaryotic Cells to Reveal Spatiotemporal Dynamics of Gene Expression. *Annual review of biochemistry* **2020,** *89*, 159-187.
250. Fu, Y.; Rocha, P. P.; Luo, V. M.; Raviram, R.; Deng, Y.; Mazzoni, E. O.; Skok, J. A., CRISPR-dCas9 and sgRNA scaffolds enable dual-colour live imaging of satellite sequences and repeat-enriched individual loci. *Nature Communications* **2016,** *7* (1), 11707.
251. Weller, R. L.; Rajski, S. R., Design, synthesis, and preliminary biological evaluation of a DNA methyltransferase-directed alkylating agent. *Chembiochem* **2006,** *7* (2), 243-245.
252. Deen, J.; Wang, S.; Van Snick, S.; Leen, V.; Janssen, K.; Hofkens, J.; Neely, R. K., A general strategy for direct, enzyme-catalyzed conjugation of functional compounds to DNA. In *Nucleic Acids Research*, 2018; Vol. 46, p e64.
253. Sharim, H.; Grunwald, A.; Gabrieli, T.; Michaeli, Y.; Margalit, S.; Torchinsky, D.; Arielly, R.; Nifker, G.; Juhasz, M.; Gularek, F.; Almalvez, M.; Dufault, B.; Chandra, S. S.; Liu, A.; Bhattacharya, S.; Chen, Y. W.; Vilain, E.; Wagner, K. R.; Pevsner, J.; Reifenberger, J.; Lam, E. T.; Hastie, A. R.; Cao, H.; Barseghyan, H.; Weinhold, E.; Ebenstein, Y., Long-read single-molecule maps of the functional methylome. *Genome Research* **2019,** *29* (4), 646-656.
254. Peinado, P.; Herrera, A.; Baliñas, C.; Martín-Padrón, J.; Boyero, L.; Cuadros, M.; Coira, I. F.; Rodriguez, M. I.; Reyes-Zurita, F. J.; Rufino-Palomares, E. E.; Lupiáñez, J. A.; Medina, P. P., Chapter 6 - Long Noncoding RNAs as Cancer Biomarkers. In *Cancer and Noncoding RNAs*, Chakrabarti, D. J.; Mitra, D. S., Eds. Academic Press: Boston, 2018; Vol. 1, pp 95-114.
255. AM, K.; M, G.; M, H.; M, G.; A, R.; D, R. M.; K, T.; A, P.; BE, B.; A, v. O.; A, R.; ES, L.; JL, R., Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **2009,** *106* (28), 11667-11672.
256. MA, F.; F, M.; AM, K.; DE, W.; BG, S.; TE, M.; CE, F.; G, S. L.; PJ, K.; C, W., Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. *Nature medicine* **2008,** *14* (7), 723-730.
257. CH, L.; Y, C., Targeting long non-coding RNAs in cancers: progress and prospects. *The international journal of biochemistry & cell biology* **2013,** *45* (8), 1895-1910.
258. K, A.; YJ, P.; AM, L.; A, S.; C, O.; D, W.; A, L.; E, L.; A, R.; M, M.; H, D.; G, D.; C, H.-M.; I, G.; C, N.; C, P., Long noncoding RNA TARID directs demethylation and activation of the tumor suppressor TCF21 via GADD45A. *Molecular cell* **2014,** *55* (4), 604-614.
259. van Tilborg, A. A. G.; Kompier, L. C.; Lurkin, I.; Poort, R.; El Bouazzaoui, S.; van der Keur, K.; Zuiverloon, T.; Dyrskjot, L.; Orntoft, T. F.; Roobol, M. J.; Zwarthoff, E. C., Selection of Microsatellite Markers for Bladder Cancer Diagnosis without the Need for Corresponding Blood. In *PLoS One*, 2012; Vol. 7, p e43345.

260. Uhlén, M.; Fagerberg, L.; Hallström, B. M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, Å.; Kampf, C.; Sjöstedt, E.; Asplund, A.; Olsson, I.; Edlund, K.; Lundberg, E.; Navani, S.; Szigyarto, C. A.-K.; Odeberg, J.; Djureinovic, D.; Takanen, J. O.; Hober, S.; Alm, T.; Edqvist, P.-H.; Berling, H.; Tegel, H.; Mulder, J.; Rockberg, J.; Nilsson, P.; Schwenk, J. M.; Hamsten, M.; Feilitzen, K. v.; Forsberg, M.; Persson, L.; Johansson, F.; Zwahlen, M.; Heijne, G. v.; Nielsen, J.; Pontén, F., Tissue-based map of the human proteome. *Science* **2015,** *347* (6220), 1260419.
261. Furney, S. J.; Higgins, D. G.; Ouzounis, C. A.; López-Bigas, N., Structural and functional properties of genes involved in human cancer. *BMC Genomics* **2006,** *7* (1), 1-11.
262. Srinivasan, M.; Sedmak, D.; Jewell, S., Effect of Fixatives and Tissue Processing on the Content and Integrity of Nucleic Acids. *The American Journal of Pathology* **2002,** *161* (6), 1961-1971.
263. Hobro, A. J.; Smith, N. I., An evaluation of fixation methods: Spatial and compositional cellular changes observed by Raman imaging | Elsevier Enhanced Reader. *Vibrational Spectroscopy* **2017,** *91*, 31-45.
264. C, H.; S, O.; J, v. H.; S, D., Preservation of large-scale chromatin structure in FISH experiments. *Chromosoma* **2007,** *116* (2), 117-133.
265. Nicholson, J. M.; Cimini, D., Cancer Karyotypes: Survival of the Fittest. *Front Oncol* **2013,** *3*, 148.
266. Weise, A.; Heller, A.; Starke, H.; Mrasek, K.; Kuechler, A.; Pool-Zobel, B. L.; Claussen, U.; Liehr, T., Multitude multicolor chromosome banding (mMCB) - a comprehensive one-step multicolor FISH banding method. *Cytogenetic and Genome Research* **2003,** *103* (1-2), 34-39.
267. Rondón-Lagos, M.; Verdun Di Cantogno, L.; Marchiò, C.; Rangel, N.; Payan-Gomez, C.; Gugliotta, P.; Botta, C.; Bussolati, G.; Ramírez-Clavijo, S. R.; Pasini, B.; Sapino, A., Differences and homologies of chromosomal alterations within and between breast cancer cell lines: a clustering analysis. *Mol Cytogenet* **2014,** *7*, 8.
268. Imai, R.; Nozaki, T.; Tani, T.; Kaizu, K.; Hibino, K.; Ide, S.; Tamura, S.; Takahashi, K.; Shribak, M.; Maeshima, K., Density imaging of heterochromatin in live cells using orientation-independent-DIC microscopy. *Molecular Biology of the Cell* **2017,** *28* (23), 3349-3359.
269. Bessler, J. B.; Andersen, E. C.; Villeneuve, A. M., Differential Localization and Independent Acquisition of the H3K9me2 and H3K9me3 Chromatin Modifications in the Caenorhabditis elegans Adult Germ Line. *PLoS Genetics* **2010,** *6* (1), e1000830.
270. Chen, F.; Tillberg, P. W.; Boyden, E. S., Expansion microscopy. *Science* **2015,** *347* (6221), 543-548.
271. Truckenbrodt, S.; Maidorn, M.; Crzan, D.; Wildhagen, H.; Kabatas, S.; Rizzoli, S. O., X10 expansion microscopy enables 25-nm resolution on conventional microscopes. *EMBO Reports* **2018,** *19* (9), e45836.
272. Chozinski, T. J.; Halpern, A. R.; Okawa, H.; Kim, H.-J.; Tremel, G. J.; Wong, R. O. L.; Vaughan, J. C., Expansion microscopy with conventional antibodies and fluorescent proteins. *Nature Methods* **2016,** *13* (6), 485-488.
273. Chen, F.; Wassie, A. T.; Cote, A. J.; Sinha, A.; Alon, S.; Asano, S.; Daugharthy, E. R.; Chang, J.-B.; Marblestone, A.; Church, G. M.; Raj, A.; Boyden, E. S., Nanoscale imaging of RNA with expansion microscopy. *Nature Methods* **2016,** *13* (8), 679-684.
274. Tsanov, N.; Samacoits, A.; Chouaib, R.; Traboulsi, A. M.; Gostan, T.; Weber, C.; Zimmer, C.; Zibara, K.; Walter, T.; Peter, M.; Bertrand, E.; Mueller, F., smiFISH and

FISH-quant - a flexible single RNA detection approach with super-resolution capability. *Nucleic Acids Research* **2016,** *44* (22), e165.
275. Tillberg, P. W.; Chen, F.; Piatkevich, K. D.; Zhao, Y.; Yu, C. C.; English, B. P.; Gao, L.; Martorell, A.; Suk, H. J.; Yoshida, F.; DeGennaro, E. M.; Roossien, D. H.; Gong, G.; Seneviratne, U.; Tannenbaum, S. R.; Desimone, R.; Cai, D.; Boyden, E. S., Protein-retention expansion microscopy of cells and tissues labeled using standard fluorescent proteins and antibodies. *Nature Biotechnology* **2016,** *34* (9), 987-992.
276. Ku, T.; Swaney, J.; Park, J. Y.; Albanese, A.; Murray, E.; Cho, J. H.; Park, Y. G.; Mangena, V.; Chen, J.; Chung, K., Multiplexed and scalable super-resolution imaging of three-dimensional protein localization in size-adjustable tissues. *Nature Biotechnology* **2016,** *34* (9), 973-981.
277. Hell, S. W.; Wichmann, J., Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Optics letters* **1994,** *19* (11), 780-782.
278. Jungmann, R.; Avendano, M. S.; Woehrstein, J. B.; Dai, M.; Shih, W. M.; Yin, P., Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT. *Nature Methods* **2014,** *11* (3), 313-318.
279. Gao, R.; Asano, S. M.; Boyden, E. S., Q&A: Expansion microscopy. *BMC Biology* **2017,** *15*, 50.
280. Dougherty, R., Extensions of DAMAS and Benefits and Limitations of Deconvolution in Beamforming | Aeroacoustics Conferences. In *11th AIAA/CEAS Aeroacoustics Conference*, Monterey, California, 2005.
281. TA, P.; JL, G., Ribosomal DNA and the nucleolus in the context of genome organization. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* **2019,** *27* (1-2), 109-127.
282. Liu, B.; Du, Q.; Chen, L.; Fu, G.; Li, S.; Fu, L.; Zhang, X.; Ma, C.; Bin, C., CpG methylation patterns of human mitochondrial DNA. *Scientific Reports* **2016,** *6* (1), 1-10.
283. Audugé, N.; Padilla-Parra, S.; Tramier, M.; Borghi, N.; Coppey-Moisan, M., Chromatin condensation fluctuations rather than steady-state predict chromatin accessibility. *Nucleic Acids Research* **2019,** *47* (12), 6184-6194.
284. Chen, D.; Dundr, M.; Wang, C.; Leung, A.; Lamond, A.; Misteli, T.; Huang, S., Condensed mitotic chromatin is accessible to transcription factors and chromatin structural proteins. *The Journal of cell biology* **2005,** *168* (1), 41-54.
285. Cohen, S.; Segal, D., Extrachromosomal Circular DNA in Eukaryotes: Possible Involvement in the Plasticity of Tandem Repeats. *Cytogenetic and Genome Research* **2009,** *124* (3-4), 327-338.
286. Cohen, S.; Agmon, N.; Sobol, O.; Segal, D., Extrachromosomal circles of satellite repeats and 5S ribosomal DNA in human cells. *Mobile DNA* **2010,** *1*, 11.
287. Shibata, Y.; Kumar, P.; Layer, R.; Willcox, S.; Gagan, J. R.; Griffith, J. D.; Dutta, A., Extrachromosomal microDNAs and chromosomal microdeletions in normal tissues. *Science* **2012,** *336* (6077), 82-86.
288. Turner, K. M.; Deshpande, V.; Beyter, D.; Koga, T.; Rusert, J.; Lee, C.; Li, B.; Arden, K.; Ren, B.; Nathanson, D. A.; Kornblum, H. I.; Taylor, M. D.; Kaushal, S.; Cavenee, W. K.; Wechsler-Reya, R.; Furnari, F. B.; Vandenberg, S. R.; Rao, P. N.; Wahl, G. M.; Bafna, V.; Mischel, P. S., Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nature* **2017,** *543* (7643), 122-125.

289. SH, K.; PG, M.; MK, L.; EM, S.; LA, P.; T, M., Spatial genome organization during T-cell differentiation. *Cytogenetic and genome research* **2004,** *105* (2-4), 292-301.
290. Faulkner, E. L.; Thomas, S. G.; Neely, R. K., An introduction to the methodology of expansion microscopy. *The International Journal of Biochemistry & Cell Biology* **2020,** *124*, 105764.
291. AT, W.; Y, Z.; ES, B., Expansion Microscopy: Principles and Uses in Biological Research. *Nature methods* **2019,** *16* (1), 33-41.
292. Castro, R.; Rivera, I.; Martins, C.; Struys, E. A.; Jansen, E. E. W.; Clode, N.; Graça, L. M.; Blom, H. J.; Jakobs, C.; Almeida, I. T. d., Intracellular S - adenosylhomocysteine increased levels are associated with DNA hypomethylation in HUVEC. *Journal of Molecular Medicine* **2005,** *83* (10), 831-836.
293. Huang, L.-K.; Wang, M.-J. J., Image thresholding by minimizing the measures of fuzziness. *Pattern recognition* **1995,** *28* (1), 41-51.
294. Berg, S.; Kutra, D.; Kroeger, T.; Straehle, C. N.; Kausler, B. X.; Haubold, C.; Schiegg, M.; Ales, J.; Beier, T.; Rudy, M.; Eren, K.; Cervantes, J. I.; Xu, B.; Beuttenmueller, F.; Wolny, A.; Zhang, C.; Koethe, U.; Hamprecht, F. A.; Kreshuk, A., ilastik: interactive machine learning for (bio)image analysis. *Nature Methods* **2019,** *16* (12), 1226-1232.