Summer 2016

Using Zinc Finger Proteins as a Diagnostic Tool for the Detection of a Cancer Biomarker

Anu Kini
Western Kentucky University, anu.kini355@topper.wku.edu

USING ZINC FINGER PROTEINS AS A DIAGNOSTIC TOOL FOR THE DETECTION OF A CANCER BIOMARKER

A Thesis
Presented to
The Faculty of the Department of Chemistry
Western Kentucky University
Bowling Green, Kentucky

In Partial Fulfillment
Of the Requirements for the Degree
Master of Science

By
Anu Ganesh Kini

August 2016

USING ZINC FINGER PROTEINS AS A DIAGNOSTIC TOOL FOR THE DETECTION OF A CANCER BIOMARKER

Date Recommended 5/18/16

Dr. Moon-Soo Kim, Supervisor of Thesis

Dr. Kevin Williams

Dr. Blairanne Williams

5/18/16

Dean, Graduate Studies and Research Date

ACKNOWLEDGEMENTS

I would like to thank Dr. Moon-Soo Kim, my research advisor for supervising my project. I appreciate her efforts in providing me with her directions and instructions. Her support and training will definitely benefit me in the future.

I would also like to thank my committee members Dr. Kevin Williams and Dr. Blairanne Williams for reading and providing their valuable input to my thesis.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

USING ZINC FINGER PROTEINS AS A DIAGNOSTIC TOOL FOR THE DETECTION OF A CANCER BIOMARKER

Anu Kini August 2016 52 Pages

Directed By: Dr. Moon-Soo Kim, Dr. Kevin Williams, Dr. Blairanne Williams

Department of Chemistry Western Kentucky University

RASSF1A is a tumor suppressor gene which loses its function due to methylation of CpG islands on its promoter region. Detection of methylation leads to early diagnosis of cancer.

Zinc finger proteins are capable of detecting a specific DNA sequence and Methyl binding domain can bind to the methyl group on the CpG, using this idea mCpG SEER- Lac system makes use of a split protein, β-lactamase. Lac A attached to the ZFP and Lac B attached to the MBD protein. On binding to the DNA, the Lac A and Lac B come in close proximity with each other causing a reassembly and activation of the enzyme. In the presence of a substrate, the activated β-lactamse enzyme hydrolyzes the β-lactam bond in the substrate and shows a color change from yellow to red in the presence of a methylated cognate DNA.

The study suggests that a solution based assay was not as specific in differentiating signal intensities between methylated and non-methylated DNA. It was also not sensitive in measuring dose dependent signals. Zinc finger array could successfully show relatively low signals for non-methylated DNA. The findings of the study show that MBD2 shows higher preference for mCpG than MBD1 in the mCpG SEER-Lac system and oligonucleotides with a 2 bp spacing between methylation and ZF target site shows higher signals than the 3 bp spacing. Due to it's specificity and sensitivity, it serves as a potential diagnostic tool to detect cancer.

CHAPTER 1

LITERATURE AND INTRODUCTION

1. RASSF1A gene

Carcinogenesis is a result of malfunction in various processes involving different regulators and factors. Proteins encoded by tumor suppressor gene (TSG) are involved in inhibition of cell survival and growth. Hence, silencing of a TSG due to mutation can result in the inhibition of these negative regulatory proteins. Studying these genes may provide more information about the cancer and help in early diagnosis. TSGs regulate the rate of proliferation and apoptosis (1). TSGs prevent proto-oncogenes from turning into oncogenes due to any mutagenesis.

Ras association domain family (RASSF) is a TSG and a part of the Ras family proteins, which are ubiquitously expressed in all cell lines (2,3). Ras proteins in normal conditions are associated with oncogenic effects such as decreased cell death and increased cell proliferation (3). Epigenetic modifications cause changes in the RASSF1A gene due to external factors without changing the DNA sequence. Epigenetic inactivation of RASSF1A due to a process where unusual number of CG dinucleotides get methylted, called hypermethyaltion on the promoter region causes inactivation of the gene (4). Silencing of RASSF1A may be correlated with activation of Ras effectors in the Ras-dependent tumors. The major Ras effectors are K-ras, H-ras and N-ras (5).

Currently, the RASSF consists of 10 members. The functions of few are known distinctly. RASSF1 is a member of six related proteins and contain multiple splice variants. Each protein contains Ras associated domain and a C-terminal SARAH (sav/RASSF/Hpo) protein-protein interaction (5). Another member of the RASSF family, RASSF2 plays a major role in cell apoptosis and it is also seen to be down

regulated in tumors due to methylation, histone deacetylation and deletion of the gene. It was proven that reintroduction after deletion hindered and deletion enhanced the process of tumor genesis (6). RASSF3 can bind to Ras and inhibit cell growth, but it is not downregulated in tumors. The involvement of RASSF3 in cancer remains unknown (7). RASSF4 also binds to Ras and induces cell apoptosis (4). RASSF5 was initially called Nore1a and was the first family to be cloned (8). It is pro-apoptotic and stops cell growth through the Ras mediated pathway (9). It is very often inactivated in human tumors due to promoter methylation. It may regulate the pro-apoptotic kinase MST1 in the Ras mediated pathway (10). RASSF5 can also form a complex with the Ras related protein (Rap). which regulates lymphocyte adhesion and acts as a TSG (11). RASSF6 possesses similar biological properties. It is frequently downregulated in tumors (12) and its location is suggested to be involved in bronchiolitis inflected by respiratory syncytial virus (13). Two other identified Ras association domain proteins are RASSF7 and RASSF8. They do not display homology with RASSF1 through 6 (14).

Table 1: Summary of Ras association domain family members and their functions (4,6,7,12)

Member	Function
RASSF1 (Isoforms A-G)	Activation of cell death, cell cycle, and microtubule formation
RASSF2	Apoptosis via cell cycle arrest
RASSF3	Binds to Ras to inhibit cell growth
RASSF5	Pro-apoptotic and kills cells in a Ras dependent manner
RASSF6	Promotes apoptosis to induce cell death and tumor cell survival.

1.1 Locus and Gene description

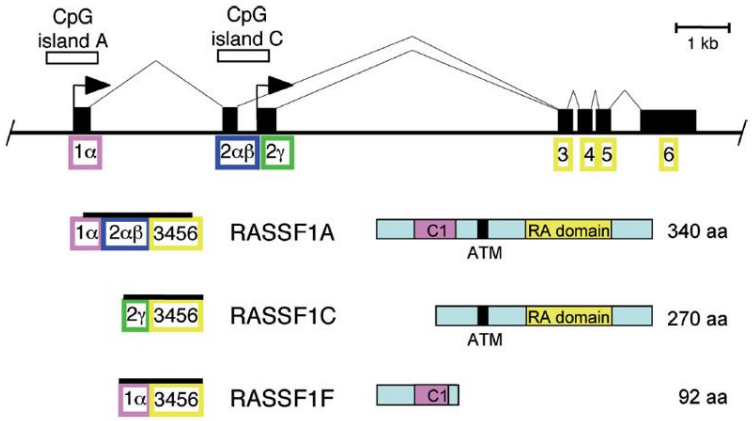

Figure 1: The promoters of RASSF1 are located in the CpG islands (hollow rectangles). Three major isoforms are RASSF1A, RASSF1C and RASSF1F and encompass a different combination of the exons 3,4,5 and, 6 remain common for all isoforms (15).

The gene locus of RASSF1 spans about 11,151 bp of the human genome. It contains eight different exons and alternative splicing gives rise to two different promoters 3.5 kb apart (16). Alternative splicing yields 7 transcripts (A-G). Isoforms A and C are ubiquitously expressed and isoform B is expressed in the cells of hemopoietic system. Isoform D and E are produced in cardiac and pancreatic cells respectively. There are two CpG islands associated with RASSF1 promoters. The smaller promoter of the two comprises of 737 bp which encompasses 85 CpGs and is observed in RASSF1D, RASSF1E, RASSF1F and RASSF1G. The second larger CpG island spans 1,365 bp and encompasses 139 CpGs which is encoded in RASSF1B and RASSF1C (16).

RASSF1A encompasses the smaller CpG island 1α and spans one of the two exons of the larger CpG island 2 αβ and the last four exons remain common for all isoforms. The Ras association domain encoded by exons 4 and 5 defines the RASSF1

gene and is located at the C-terminus of isoforms A-E. This domain is involved in interactions of Ras and other GTPases. The Ras association domain is composed of five-stranded mixed β-sheets and three α-helices. These domains dimerize by intermolecular disulfide bonds formed between the cysteine residues from each monomer. RASSF1A and RASSF1D-G have a N-terminal protein kinase conserved region called C1 domain encoded by exons 1 α and 1 β. This is the diacylglycerol binding domain (DAG) that is responsible for regulating activity of kinase. Overlapping the C1 region, is the zinc finger binding domain ZNF-NFX. A SARAH domain (sav/RASSF/Hpo) is present at the C-terminus of RASSF1A-E. These domains mediate interactions between proteins for Hpo/Sav homotypic interactions with MST1 (16).

1.2 Biological role of RASSF1A

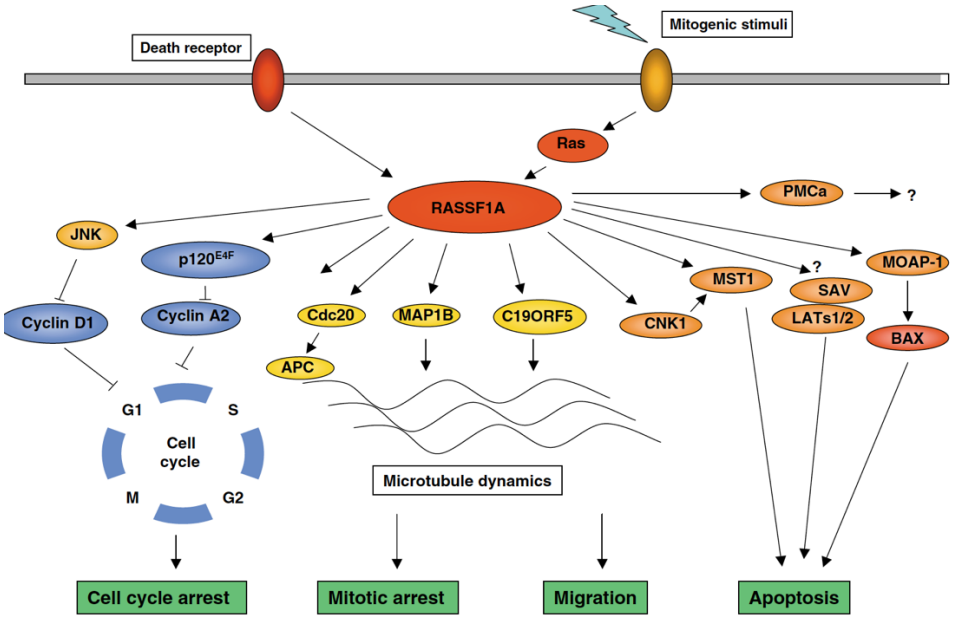

Figure 2: Pathways used by RASSF1A in different biological functions (17).

1.2.1 Deregulating microtubule dynamics

Polymers of tubulin, called as microtubules are capable of assembling and disassembling at various stages of cell cycle. Their function is to ensure segregation of sister chromatid and regulate cytokinesis (18). Microtubule associated protein (MAP) binds directly to tubulin. RASSF1A could be associated with microtubules by mediating MAPs. MAP1b promotes polymerization of microtubules. Microtubules can switch between stages of elongation and shortening; this process is called as dynamic instability. Stabilizing the microtubule causes loss of its function. RASSF1A is localized to the microtubules and promotes its stabilization (5).

1.2.2 Cyclin dependent deregulation of cell cycle

c-Jun-NH2-Kinase is also called stress-activate protein kinase (SAPK). It has the ability to phosphorylate the N-terminal of transcription factor, c-Jun, activating it as a response to various stress. JNK is involved in multiple pathways and is an important mediator in the process of cell death. One of its many pathways involve Ras effectors proteins. Overexpression of RASSF1A suppresses Ras induced phosphorylation of (JNK) c-Jun-NH2-Kinase, which in turn decreases the levels of cyclin D1(19). Cyclin D1 is an essential factor required by the cells for progression of cell cycle through the Rb checkpoint . RASSF1A induced decrease in JNK signaling and cyclin D1 levels causes cell cycle arrest in the G1 phase of the cycle (20).

1.2.3 Cell apoptosis

Mammalian sterile 20 like kinase, MST1 and MST2 are pro-apoptotic serine and threonine kinases that activate JNK pathway in a way discussed earlier. MST1 helps RASSF1A to phosphorylate histone, which causes condensation of chromatin and apoptosis due to lack of transcriptional control (5).

A scaffold protein called CNK1 augments RASSF1A induced cell death. This is a connector enhancer of KSR (CNK) and a c-Raf1 binding protein which is a mediator of Ras induced Raf activation. CNK1- RASSF1A surpasses the growth cancer cells via MST1 leading to apoptosis by the pro-apoptotic MST1 pathway (21).

Modulator of apoptosis (MOAP1) is another pro-apoptotic protein, it get's activated by stimulation of death receptors and arms (tumor necrosis factor receptor) TNFR-MOADP1 complex, localizing RASSF1A to the death receptor (22). This complex also activates another pre-apoptotic protein called bcl-2-like protein 4 (Bax) and p21, which is called cyclin dependent kinase inhibitor 1. RASSF1A-MOAP1 initiates a conformational change in MOAP1, exposing its hidden BH3-like domain and enabling it to interact with Bax. In this way, Bax inserts itself into the mitochondrial membrane and executes cell apoptosis. This interaction is dramatically amplified by activated K-Ras and absence of RASSF1A intervenes with the ability of K-Ras to activate Bax. Thus a Ras-RASSF1A-MOAP1 complex is important for Ras induced cell apoptosis (22).

1.3 Methylation of RASSF1A

DNA methyltransferase (DNMT) is an enzyme responsible for modifying one position in the target DNA sequence. In humans, the DNMT falsely recognizes the DNA sequence and methylates the cytosine in CpG rich areas. Sequences that may get methylated are exons, introns, repetitive regions in the DNA, regulatory elements, and other transposons (23). DNMT catalyzes the transfer of a methyl group from a S-adenosyl-L-methionine to cytosine of a hemimethylated DNA (24). Promoter based aberrant hypermethylation of tumor suppressor genes along with hypomethylation is associated with high rates of mutagenesis and it is an essential factor in the process of carcinogenesis. This suggests that methylation of DNA is an important function in the genetic integrity system (25). When DNMT methylates a DNA sequence, it recruits histone deacetylases from the neighboring region of the DNA, repressing transcription of the gene (26).

Gene silencing of tumor suppressor gene RASSF1A, is a fundamental epigenetic inactivation mechanism of genes related to cancer (27). In case of RASSF1A, promoter hypermethylation is a most common inactivation process which causes loss of functionality of the gene (21). Aberrant tumor hypermethylation of RASSF1A was detected in several tumors like breast, lungs, colon, kidney, prostate and bladder cancer (28). The cytosine on the CpG islands in the promoter region of the gene is methylated on an account of epigenetic inactivation. This causes failure of RASSF1A gene to disrupt the cell cycle allowing uncontrolled growth of cells. Detection of methylated DNA was correlated to availability of methylated tumor-derived DNA and detection of non-methylated DNA suggested absence of cancer (29). Unlike proteins, DNA may be amplified to a certain extent, this fact gave researchers an advantage to detect methylated DNA. For a long period of time,

methylated tumor-derived DNA has been known to be available in the blood stream (30). It was later studied that methylated RASSF1A was also found in other body fluids corresponding to the tumor of that tissue (30,31)

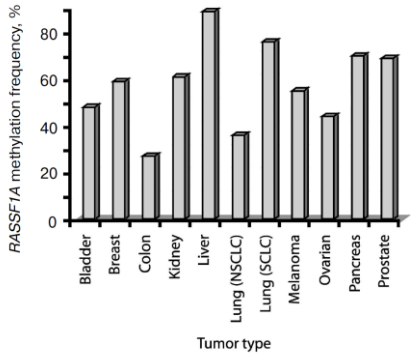

Figure 3: A bar graph indicating percentage frequency of methylation of RASSF1A in different tumors (21).

A study conducted by Yihao et al. suggests that RASSF1A gets hypermethylated in a CG rich region (32). They showed that methylated RASSF1A lacks a TATA box CG region which is essential for transcription initiation by the RNA polymerase 2 complex. This was consistent over short distances like 250 bp, and was observed to be a common trait of active promoters. The methylated DNA after treatment with bisulfite facilitates the conversion of methyl cytosine to uracil, which is amplified using real-time PCR and sequenced. Their results in bisulfite sequencing of RASSF1A CpG islands showed 78.5% of the GC sites which were methylated in testicular tumor (30).

Some studies conducted by Wei Meng suggests that RASSF2A also get hypermethylated in carcinomas (14). And it concluded that RASSF1A was methylated in 68% and RASSF2A in 7% of transitional cell carcinoma. Hence it can be concluded that RASSF1A is a more dependable biomarker than RASSF2A. This

study indicated that detection of promoter methylation in urine specimen together showed good sensitivity (77%) and high specify (94%) identifying bladder cancer tissues (14).

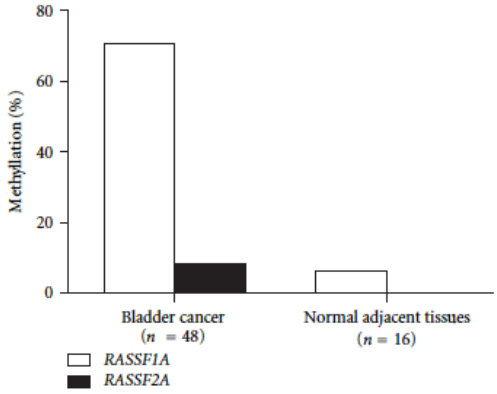

Figure 4: Methylation in RASSF1A as compared to RASSF2A (32).

2. DNA Binding domain

DNA-binding Proteins can “read” nucleotide sequences because of two suitable conditions (33). First being the B form of DNA, present in most cell population, has base pairing from one phosphate backbone to another. The number of hydrogen bonds and the final structure that forms due to the hydrogen bonds (A forms two bond with T and G forms a three bond with C) helps DNA-binding proteins to distinguish between different nucleotide bases. The methyl group on the thymine is a sequence recognition element that encourages the DNA recognition process. Second being the ability to accommodate the DNA binding protein. The major groove of the DNA is larger in area than the minor groove and can easily accommodate the helical structure of the protein. Hence, the nucleotides on the exterior face of the helix should be able to read and recognize target nucleotide sequences without disruption in the sequence of the DNA’s initial structure.

2.1 Zinc finger proteins (ZFPs)

The Cys_2-His_2 is the most common type of zinc finger (ZF) and is the first member of the extensively expanding group of the zinc-binding modules. They are one of the most abundantly expressed proteins in the eukaryotic cell (34). The function of a ZFwas reported in 1989, since then the structures of all its complexes have been discovered starting with Zif268. Wright et. al (35) defined ZFs to be small, functional, independently folded domain coordinated with one or more zinc molecules in order the structure. The most frequently used is the one where the amino acids form a ββα framework (36). These folds are stabilized by the hydrophobic coordination network of the Zn and this framework provides an insight to how the ZFP interact with the DNA. Twenty-five of the 30 amino acids

in the repeat, fold around the zinc to form a ‘finger’ and the rest of the five amino acids serve as linkers between consecutive fingers. The zinc ion forms a tetrahedral complex coordinated to two pairs of cysteine and histidine. This hydrophobic interaction of the coordination network between the Zn and the cysteine and histidine residues stabilize the fingers (37). The amino acids in each ZF have the affinity with specific nucleotides. Therefore, every finger is capable of specifically and selectively recognizing 3-4 nucleotides of the DNA (34). Multiple ZF can be arranged into tandem arrays and can recognize a set of nucleotides on the DNA. Using methods like combinatorial mutagenesis, where in, a large number of mutants are screened for DNA specificity (38), it is possible to create modules of six zinc fingers that can potentially recognize 18 bp. The length of 18 bp would be long enough for ZFPs to recognize any desired DNA sequence in the human genome. In Zif268 a canonical type of screening to recognize the nucleotide sequence was observed (34). The amino acids in positions -1, 3, 6 contact the 3’, middle, 5’ nucleotides respectively. Different ZF DNA binding proteins may be engineered to bind to unique DNA sequnces (39).

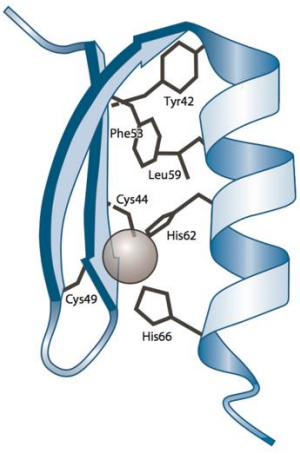

Figure 5 Structure of a ZFP showing zinc ion in the center (grey circle) forming co-ordinate bonds with two Cys and two His residues, showing ββα framework. (34)

In principle, multi-finger proteins in modular assembly, can be constructed by assembling predefined ZF modules in any order to recognize any desired DNA sequences in a modular fashion (40). A scarcely occurring hindrance in recognition pattern for modular assembly is called “context-dependent effect” (40). For example, due to this effect, if the amino acid on the α-helix is an Aspartate, then it recognizes the guanine or thymine on the subsite of the previous ZF, when they happen to be the first nucleotide in the subsite. This is due to the helical nature of the amino acid and is called target site overlap (TSO). Other types of context-dependent effects may be the result of sequence irregularities in the DNA structure due to base stacking. Recognition of this effect led to the development of second generation ZFPs that would have higher selectivity for longer sequences (40).

2.2 Modular assembly and construction of ZFPs

Each ZF recognizes 3 bp of DNA, which seems to be simple mode of DNA recognition. This may inspire creation of multi-finger proteins that can recognize any extended DNA sequences. Multiple approaches have been taken to identify optimized individual ZF modules to recognize one of the 64 possible 3 bp DNA subsites, using combination of rational design and selection (40). However, there are other ways to assemble ZFs. For example, Carroll et. al (41) used a PCR based method to construct multi-finger proteins. Wright et. al (42) developed an archive of plasmids encoding more that 140 ZF modules along with a complementary web-based software called ZiFiT to verify potential ZF targets on the gene of interest.

The Barbas and ToolGen domains are the two most commonly used sets of modular assembly fingers. Both domains cover 3 bp GNN, most ANN, many

CNN and some TNN triplets (where N can be any of the four nucleotides). Both have a different set of fingers which allows searching and coding different ZF modules as needed (40). The main advantage of this approach is that there is no strict order to assemble the fingers and no selection step. This makes modular assembly approach fast and less complicated. The context-dependent effect is the only demerit associated with modular assembly.

A set of naturally occurring ZFs derived from human transcription factor was found to recognize different triplet subsites (39). The DNA binding specificity was determined using a yeast-based one-hybrid reporter molecule (42).

3. Methyl Binding domain

Methyl binding domains (MBDs) are proteins that are capable of interpreting DNA methylation signals of CpG. The structure of MBD in the complex is known to fold into an α/β sandwich structure which is composed of a layer of twisted β-sheets backed by another layer of α-helix and finally a hairpin loop at the C-terminus (25). Methyl groups on the DNA are recognized by making major groove contacts by five MBD residues Val-20, Arg-22, Tyr- 32, Arg-44 and Ser-45 (25). NMR studies and X-Ray images showed that the Arg-22 and Arg-44 are mainly involved in forming hydrogen bonds with the guanine in the major grove (23). This pushes the guanine residue towards the minor grove reducing the stacking interaction between the methyl cytosine and the guanine. Agr residues also form a cationic π-interaction with the cytosine (43). The other residues are also involved in forming a hydrophobic patch that helps in making contact with the mCG. Further specificity is provided by Try-34 residue owing to its hydroxyl side chain that is capable of accepting a hydrogen bond from the 4-amino group of the mC (43). This was also proved by point mutation analysis by changing the Tyr-34 with Phe or Ala (43). Replacement of Tyr-34 after mutation did not show recognition of mCpG, indicating that Tyr-34 is important for DNA binding. Mutation on Ser-45 also showed reduced DNA binding (43). Hence, it could be concluded that the DNA binding to mCpG was a result of hydrophobic and polar interactions of the five MBD residues on the major groove of methylated DNA. The hairpin loop forms contacts with the DNA phosphodiester backbone which favorably donates hydrogen bonds to anchor to the DNA at certain amino acid residues (44).

The different methyl binding domains that are known to exist are MBD1, MBD2, MBD4 and MeCP2 (23,45,46). However, all may not have methyl binding function

in the mammalian system (47). MBD2 and MBD3 have an identical structure, differing in the sizes of their introns and encode 70% of identical proteins. Yet, mammalian MBD3 does not have methyl binding properties in humans (47). MBD2 is a component of methyl-CpG binding protein 1 (MeCP1) complex. MeCP1 represses transcription in a methylation density dependent manner (48). MBD1 contains a very strong repression domain that is associated with recruitment of histone deacetylases (23). Interestingly, the function of MBD4 is to minimize mutations by detecting mismatches on the methyl-CpG in the genome, and like others it does not have a repression domain (49).

4. The mCpG SEER-LAC-Zinc Finger protein system

There are two main classes of engineered DNA-binding domains (DBD) - ZF based DBD and transcription activator-like effector (TALE) DBD. The ZFP is designed in such a way that it is capable of recognizing a specific DNA sequence. The mCpG-SEER Lac system provides enzymatic amplification of a visual signal to confirm the presence of the desired DNA sequence (36).

The SEER-Lac system depends on enzyme reassembly to recognize DNA sequences. This system consists of a split protein system consisting of an enzyme from TEM-1 β-lactamase and dissected into two halves, which are initially inactive, but reassemble when they are brought in close proximity to form an active complex in presence of a cognate DNA sequence. In nitrocefin assay, β-lactamase enzyme hydrolyzes the β-lactam ring in the substrate, nitrocefin. When one of the spilt protein is linked to a MBD protein and other to a ZF, upon activation of SEER-Lac system, a desired methylated DNA sequence can be recognized (50). MBDs or mCpG site enable the detection of aberrant methylation of tumor suppressor gene associated promoter region on RASSF1A gene (51).

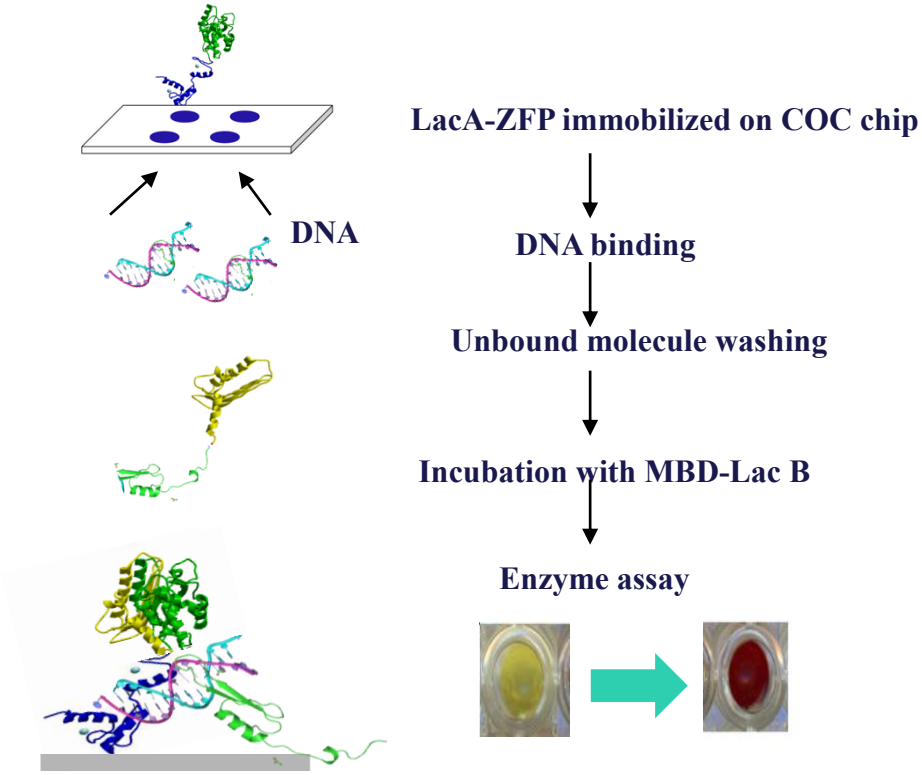

Figure 6: mCpG-SEER-Lac system for ZFp array.

5. Project Objectives

Since methylation occurs at a very primitive stage of cancer, early diagnosis of methylation of the gene will help in disease management. With this technique, we can develop a point of care testing that can easily detect the presence of methylated DNA from the urine sample of a bladder cancer patient within 3-4 hours. It may be useful to detect other carcinomas from the body fluids of the respective organ. This obviates the need of a biopsy which is penetrative in nature and painful. Our initial goal was to compare binding of target and non-target DNA using solution based assay. However, a solution based assay showed less efficiency in differentiating the signals of target and non-target due to the feasibility of the reaction provided by non-specific binding to the DNA in case of some ZFPs. Hence, we used a ZFP array which allows washing steps to wash away unbound proteins and DNA molecules from the cyclic olefin copolymer chip (COC).

CHAPTER 2

MATERIAL AND METHODS

1. Construction and expression of RASSF1A and MBD SEER proteins

All zinc finger proteins (ZFPs) were engineered using modular assembly from Barbas domain in an Sp1framework. DNA recognizing regions for each ZFP were commercially synthesized using BioBasic. ZFP (RASSF1A 150, RASSF1A 275, RASSF1A 407, RASSF1A 686) were sub-cloned between XmaI and AgeI sites of pMAL-c2X-Lac A stx2 268 by replacing the C terminal ZFP stx2-233-268. MBD2 and MBD1 was cloned between BamHI and AgeI sites of the pMaL-c2X stx2-233-Lac B replacing the N-terminal ZFP with the MBD 2 or MBD1 domain. After the plasmids were obtained, they were transformed into competent *E.coli* cells. DNA from the cells were purified using the miniprep kit (Qaigen). Plasmid DNA concentrations were measured using a Nanodrop (Thermoscientific). Approximately 3000 ng of plasmid DNA was used for restriction double digestion using XmaI and AgeI and cut smart buffer. Digested DNA was analyzed on an agarose gel.

Desired vectors of pMAL-Lac A and pMAL Lac B (6.7 kb) and RASSF1A (540 bp) along with MBD1/2 (~207/231 bp) were isolated by cutting out from the agarose gel, respectively and the DNA was extracted using gel extraction kit (Qaigen). Vector and insert were ligated using T4 DNA ligase. However, pUC-SP MBD2 first had Lac B inserted into it followed by insertion of MBD2-Lac B into pMAL vectors. The new ligated plasmid was transformed in to 10β *E.coli* competent cells which were efficient in replicating. These cells were verified for a size using a colony PCR. The same set of colonies were also grown on a separate master plate and the LacA-RASSF1A and MBD-Lac B were run in an agarose gel. Positive colonies were selected to be transformed into protein expressing BL21 competent cells for protein expression.

Table 2: Sequences of zinc fingers and their subsites.

ZFP	Finger 1	Finger 2	Finger 3	Finger 4	Finger 5	Finger 6
Target site RASSF1A 150	AAG RKDNLKN	GGC DPGHLVR	AAG RKDNLKN	AGG RSDHLTN	GGA DPGHLVR	GAG RSDNLVR
Target site RASSF1A 407	GGG RSDKLVR	GCT TSGELVR	CGC HTGHLLE	GGG RSDKLVR	GCT TSGELVR	CCC SKKHLAE

2. Cloning details

2.1 Miniprep- Plasmid isolation was achieved using a miniprep kit (Qaigen). Cells were grown overnight in 5 mL LB broth/5 mM ampicillin. Cells were centrifuged for 15 min at 4000 rmp at 4°C. Supplier instruction provided by the kit was followed. Exactly 30 μL lukewarm water was added to the column and allowed to sit at room temperature for 5 min. This is to allow the DNA to dissolve in the water. The tubes were centrifuged for 1 minute at 13,000 rpm at 4°C and checked for the concentration of the DNA.

2.2 Restriction digestion: 3000-3300 ng of all plasmid DNA were digested with AgeI and XmaI (Lac A-RASSF1A). This was carried out in duplicates or quadruplicates of the same sample. Similar amounts of MBD plasmid DNA were digested with HindIII and AgeI. Double digestion was performed using high fidelity restriction enzymes at 37°C for 3 hours.

2.3 Gel separation: Digested DNAs of the zinc finger proteins were run on a 1.2% agarose gel (0.6 g agarose in 55 mL 1x TAE buffer- 40 nM Tris pH

7.6, 20 nM acetic acid and 1 mM EDTA). Ethidium bromide was added to the gel when gel solution becomes less warm. Restricted MBD DNAs were run on a 1.8% gel. RASSF1A were cut ~540bp and MBD were cut out at ~200-240 bp.

2.4 Gel extraction: The DNA gels were cut at their expected size and extracted using the gel extraction kit (Qaigen). Supplier instrcutions were followed to extract DNA from the gel. Exactly 22 μL of lukewarm water was added to the column. After 5 minutes, it was centrifuged for 1min at 13,000 rpm at 4°C.

2.5 DNA Ligation: All extracted DNA are to be plugged into a vector, pMAL-Lac A or pMAL-Lac B. All ZF were to be inserted into pMAL-Lac A-Insert and both MBD need to be inserted into pMAL-Insert-Lac B. NEB T4 DNA ligase was used for ligation. The following vector and insert ratios were used.

2.6 Transformation: New ligated plasmids were transformed into competent NEB 10β *E.coli* cells. Competent cells were thawed at room temp on ice for 10 min. Approximately 180 ng of plasmid DNA was added to the cells. It was set on ice for 20 minutes followed by a heat shock treatment at 42° C for exactly 30 seconds and placed in ice immediately for 1 min. 350 μL of SOC medium of room temperature was added and the tubes were allowed to shake for 1 hour at 37 °C. This culture was spread on LB/ampicillin plates. Plates were incubated at 37 °C overnight for 12-16 hours.

2.7 Colony PCR: These colonies were analyzed for correct DNA insertion by PCR amplification followed by running on an agarose gel. A portion of the

colony was also dotted on a LB/Ampicillin plate (master plate). The master plate is used as a reference for the positive colonies.

2.8 Transformation for protein expression: Positive colonies were transformed into competent BL21 *E.coli* (Invitrogen) cells for protein expression as explained earlier.

2.9 Protein purification: BL21 cells were cultured and induced with 0.3 M Isopropyl β-D-1-thiogalactopyranoside (IPTG) at an OD_{600} of 0.6-0.8 for 3 hours at 37˚C. Cell pellets were stored in -80˚C prior to resuspension with Zinc Buffer A (ZBA: 100 mM Tris base, 90 mM KCl, 1 mM $MgCl_2$ 100 μM $ZnCl_2$ at pH 7.5) /5 mM dithiothreitol (DTT) and 50ug/RNAse A. Cells were sonicated for 1 minute to lyse cells and was centrifuged to collect the lysate. This was subjected to an amylose resin that as equilibrated with ZBA and 5 mM DTT and washed with 2 M NaCl in ZBA and 1 mM tris(2-carboxyethyl) phosphine (TCEP) and was finally eluted in five 1ml fractions of 10 mM maltose and 1 mM TCEP in ZBA. Concentrations of the purified proteins were analyzed using Bradford assay and the purity of the purified proteins was verified using a Coomassie stained sodium dodecyl polyacrylamide gel electrophoresis (SDS-PAGE) using bovine serum albumin (BSA) as standards. Purified proteins were stored on ice at 4˚C until utilized.

3. DNA oligonucleotides

DNA oligonucleotides were designed to have methyl CpG on the sense strand along with the ZFP target area on the anti-sense strand. Oligonucleotides were annealed in a thermal cycler (BioRad, Hercules, California) using the following protocol. Heat to 95˚C for 7 minutes, cool down to 56˚C and equilibrate at 56˚C for 5 minutes, cooling it further down to 25˚C and equilibrating at 25˚C for 10 minutes.

Table 3: List of oligonucleotides. Red color indicates methylatd target and purple color indicates ZFP target

Oligonucleotide Name	Sequence (5'-3')
149 mCpG-2	GCC TAmC GTA CTT GCC CTT CCT TCC CTC ACC G CGG ATGmCAT GAA CGG GAA GGA AGG GAG TGG C
149 CpG-2	GCC TAC GTA CTT GCC CTT CCT TCC CTC ACC G CGG ATG CAT GAA CGG CAA GGA AGG GAG TGG C
407 mCpG hairpin-3	GCG TAmC GTA TGC CAG CTC CCG CAG CTC GGG TTT TCC CGA GCT GCG GGA GCT GGC ATA mCGT ACG C CGC ATG CmAT ACG GTC GAG GGC GTC GAG CCC AAA AGG GCT CGA CGC CCT CGA CCG TAT GmCA TGC G
407 CpG hairpin -3	GCG TAC GTA TGC CAG CTC CCG CAG CTC GGG TTT TCC CGA GCT GCG GGA GCT GGC ATA CGT ACG C CGC ATG CAT ACG GTC GAG GGC GTC GAG CCC AAA AGG GCT CGA CGC CCT CGA CCG TAT CCA TGC G
149mCpG 3bp-3	GTG ATG AGT GGA mCGA AGC TTG CCC TTC CTT CCC TCG ACG C CAG TAC TCA CCT GmCT TCG AAC GGG AAG GAA GGG AGC TGC G
149 CpG 3bp-3	GTG ATG AGT GGA CGA AGC TTG CCC TTC CTT CCC TCG ACG C CAC TAC TCA CCT GCT TCG AAC GGG AAG GAA GGG AGC TGC G
149 mCpG 2bp-3	GTG ATG AGT GGA mCGA ACT TGC CCT TCC TTC CCT CGA CGC CAC TAC TCA CCT GmCT TGA ACG GGA AGG AAG GGA GCT GCG
149 CpG 2bp-3	GTG ATG AGT GGA CGA ACT TGC CCT TCC TTC CCT CGA CGC CAC TAC TCA CCT GCT TGA ACG GGA AGG AAG GGA GCT GCG
407 mCpG -4	AGT CTA AGC TCA mCGT TTG CCA GCT CCC GCA GCT CCC ATC TCA GAT TCG AGT GmCA AAC GGT CGA GGG CGT CGA GGG TAG
407 CpG -4	AGT CTA AGC TCA CGT TTG CCA GCT CCC GCA GCT CCC ATC TCA GAT TCG AGT GCA AAC GGT CGA GGG CGT CGA GGG TAG

4. EMSA (Electrophoretic mobility shift assay)

Biotinylated EMSA oligonucleotides of concentration 20 µM were annealed using a thermal cycler (BioRad, Hercules, California) at conditions of heating to 95°C for 3 minutes and cooling by 1°C per 50 seconds to 4°C. Equimolar concentrations of 100 µM of forward and reverse oligo were diluted with 500nM NaCl/ water. ZFP solution of desired concentrations was prepared using ZBA and 25% glycerol to the final concentration. Annealed DNA was serially diluted from 20 µM to 800 nM to 100 nM.

A pre-dilution mix was prepared using 150 mM KCl, 0.08% NP-40 detergent, 0.17 mg/mL BSA, 5 mM DTT and 0.83 nM annealed DNA. 3:2 proportion of pre-dilution mix and 25% glycerol and 5mM DTT in ZBA, respectively. Further series of dilutions of RASSF1A proteins and DNA were made from 1-10, in dilutions of 1:2 (in case of Lac A-RASSF1A 407, starting from 200nM of protein) or 3:4 (in case of Lac A-RASSF1A 150, starting from 800nM of protein). All 10 samples were run on a 10 % polyacrylamide gel at 100 mV for 50 minutes. Transfer apparatus was used for the transfer of protein-DNA bands on a nitrocellulose membrane. This process was run for 1 hour at 100 mV. The nitrocellulose membrane was then cross-linked in a UV-cross-linker for 4 minutes. Light shift EMSA kit (Pierce, Rockford, IL, USA) was used according to kit's instructions to block and wash the membrane.

5. Nitrocefin assay

A solution based assay of nitrocefin was assessed on purified proteins in a 96-well plate and on a cyclic olefin copolymer chip (COC) was used for the array. In the solution based assay, 20 μL of the RASSF1A construct was added along with equimolar concentrations of MBD1 or MBD2 construct in ZBA to make the final volume 180 μl. Different protein concentrations of 50, 125, 150, 200 and 250 nM final concentrations of proteins were used to determine a dose response curve. 20 μL of different concentrations of 300, 200, 100, 50, 5, 20, 10 and 1 nM of the oligonucleotides were added to the 96-well plate. Combinations of RASSF1A was used with MBD 1 and MBD2. After incubation for 1 hour, 20 μL of nitrocefin (Calbiochem) was added right before reading its absorbance in at 486 nM in a spectrophotometer (Molecular devices Sunnyvale, CA).

For reading ZFP array, DNA oligonucleotides were annealed using the protocol described above. ZFP immobilization and reaction area was confined using a

silicone gasket (Grace-biolab) providing an area with a diameter of 6 mm and a depth of 1mm. 5.3ul of purified MBP-Lac A-ZFP (RASSF1A) of the concentration of 2.5 μM was immobilized on the hydrophobic surface and incubated for 60-90 minutes until solution dries. Exactly 10 μl of DNA of different concentrations were added to different wells and allowed to bind for 20 minutes. The gasket was removed before washing the slide with ZBA/50 mM KCl followed by ZBA/ 0.05% Tween-20. 10 μl of 2.5 μM of purified MBD1 or 2-Lac B was added and incubated for 20 minutes to allow coupling with methylated DNA and RASSF1A protein.

CHAPTER 3

RESULTS AND DISCUSSION

1. Colony PCR results

   Gel pictures of agarose gel stained with ethidium bromide were taken using a gel imager. In figure 6, the bright spots at ~ 700 bp indicated positive colonies for MBD1-Lac A and MBD2-Lac B. Positive colonies for Lac A-RASSF1A 150, Lac- RASSF1A 275, Lac- RASSF1A 407 and Lac- RASSF1A 686 in figure 7 were observed at ~1.2 kp. Positive colonies on the master plate were used for protein expression and purification.

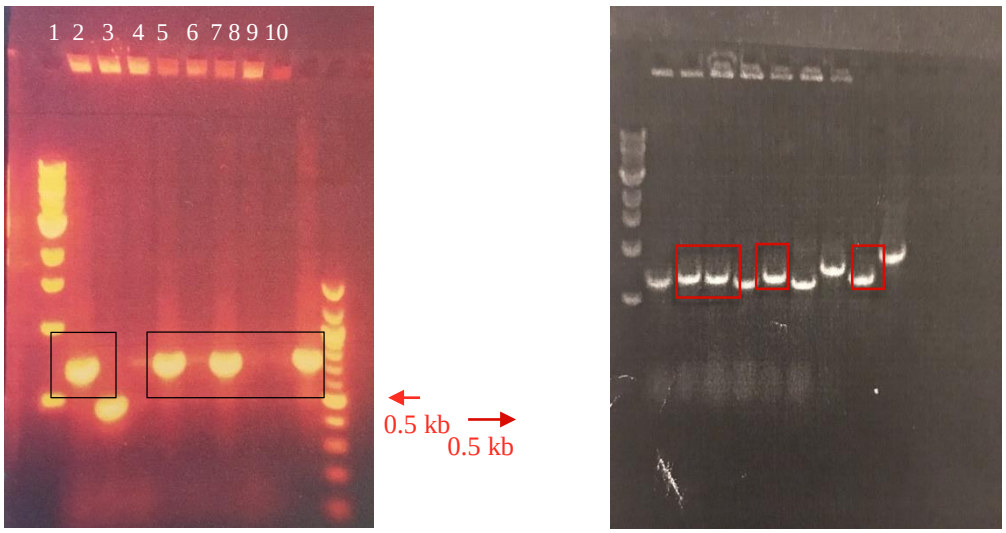

Figure 7: Gel image from colony PCR results of pMAL-MBD1-Lac B and pMAL-MBD2-Lac B. Lane 1 is the NEB 1 kb Ladder. Red/ black box indicates positive colonies at ~700 kb. Lane 10 is NEB 100 bp ladder.

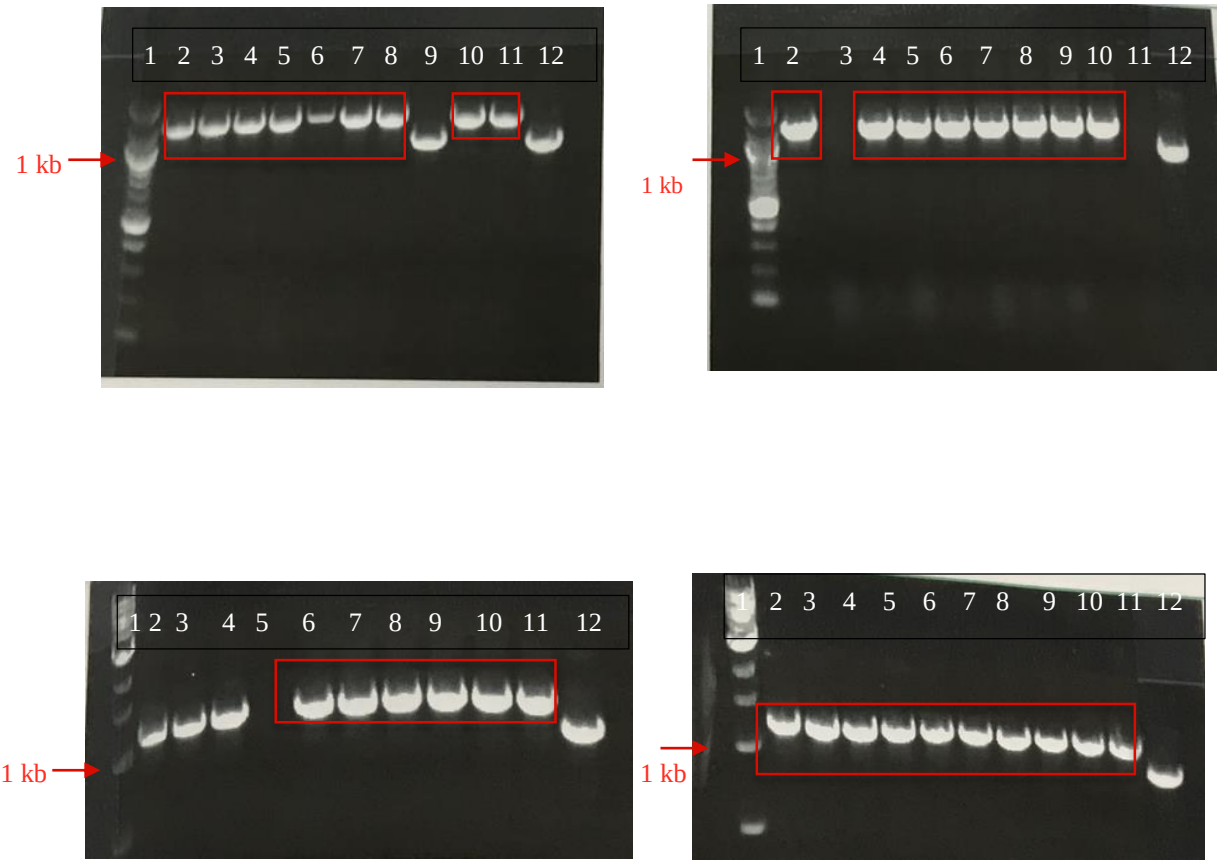

Figure 8: Gel image from colony PCR results of Lac A-RASSF1A 150, Lac A-RASSF1A 275, Lac A-RASSF1A 407 and Lac A-RASSF1A 686 (from left to right). Lane 1 is the 10 kb or 100 bp. Red box indicates positive colonies at ~1.2 kb.

2. SDS-PAGE gel

Purified proteins were subjected to SDS-PAGE to check purity of the protein on the basis of separation from other proteins. In all gels, the first lane was a protein standard for reference. Lanes 2-5 were the mixture of proteins in the washing solutions from the purification steps. Lane 6-10 were the proteins of interest eluted with a 10 mM maltose solution from the amylose resin. These lanes look free from other proteins. Both MBD1-Lac B and MBD2-Lac B were observed near 62.5 kDa and both Lac A-RASSF1A-150 and Lac A-407 were observed at around 84 kDa.

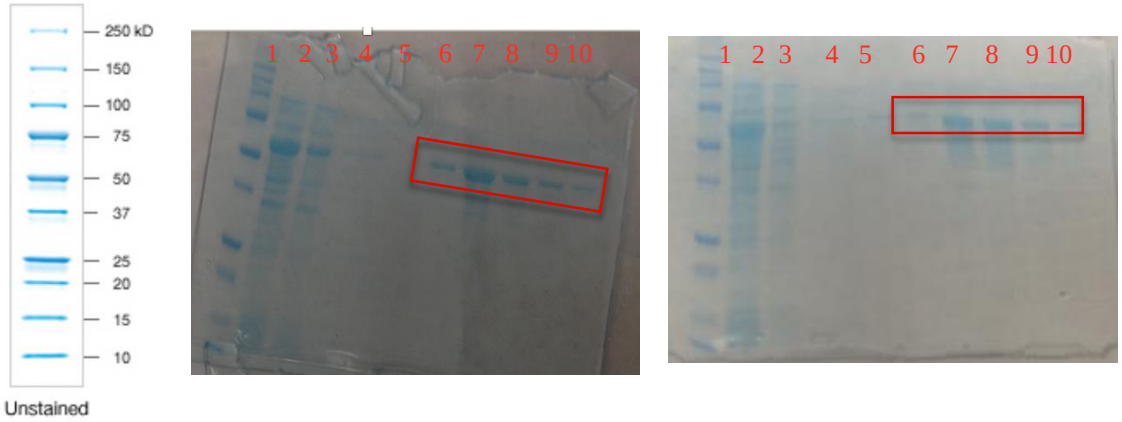

Figure 9: SDS gel picture of MBD1-Lac B and MBD2-Lac B (left to right). Red box indicates presence of protein MBD1-Lac B and MBD2-Laac B at ~62.5 kDa.

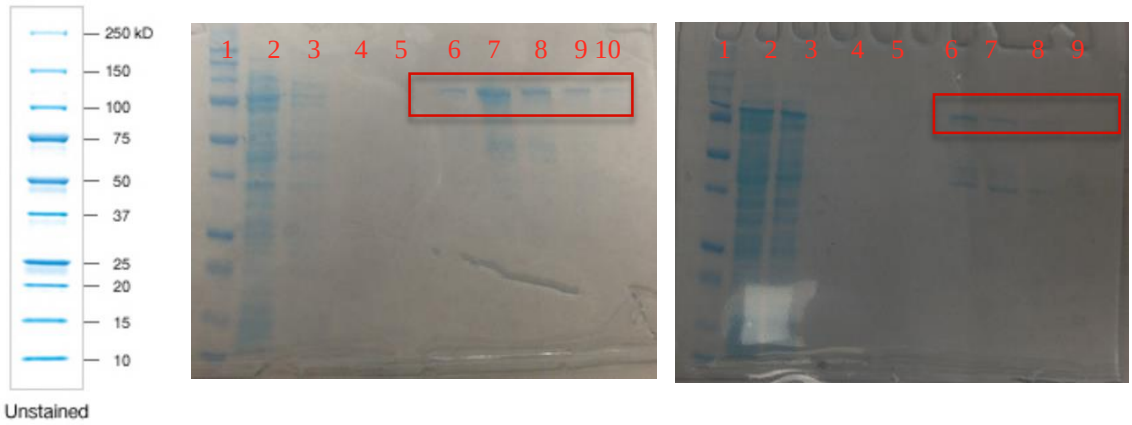

Figure 10: SDS gel picture of Lac A-RASSF1A 150 and Lac A-RASSF1A 407 (left to right). Red box indicates presence of protein Lac A-RASSF1A 150 and Lac A-RASSF1A407 at ~84 kDa.

3. Electrophoretic mobility shift assay (EMSA)

Binding affinity can be determined from the bands that yield approximately equal association and dissociation of the protein-DNA complex. Our EMSA results suggests that RASSF1A 407 shows binding affinity of 4.86 nM and binds to its cognate DNA sequence very strongly. However, the binding affinity of RASSF1A 150 was found to be 253.1 nM, indicating a weaker binder than RASSF1A 407.

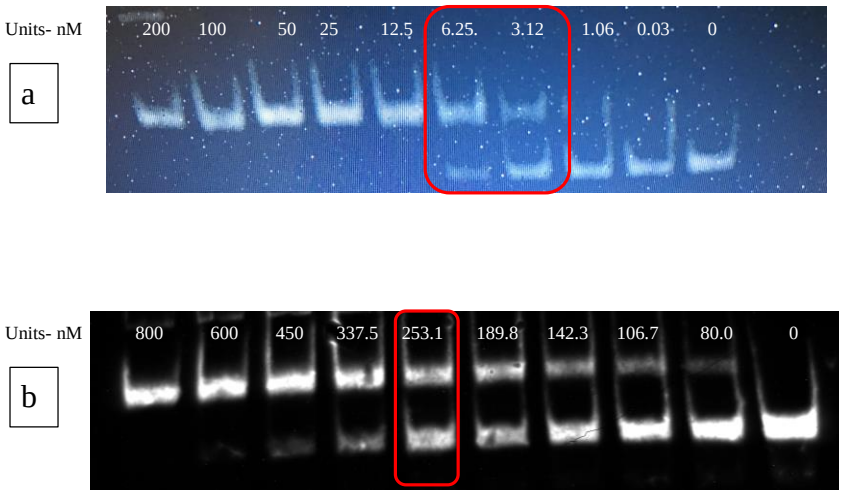

Figure 11: EMSA results of protein Lac A-RASSF1A 407 (a) and Lac A-RASSF1A 150 (b). Figures on the top indicate the protein concentration in nM and red rectangles indicate the kd value.

4. Experiments from solution based nitrocefin assay

Table 4: Sets of experiments in solution based assay

#	ZFP	MBD	Protein concentration (nM)	DNA concentration (nM)	Oligonucleotide sequence #
1	Lac A-RASSF1A 150	MBD1-Lac B	145	200,100,50,20	149 mCpG-2 149 CpG-2
2	Lac A-RASSF1A 686	MBD1-Lac B	125	100, 50	661 mCpG hairpin-2 661 CpG hairpin-2

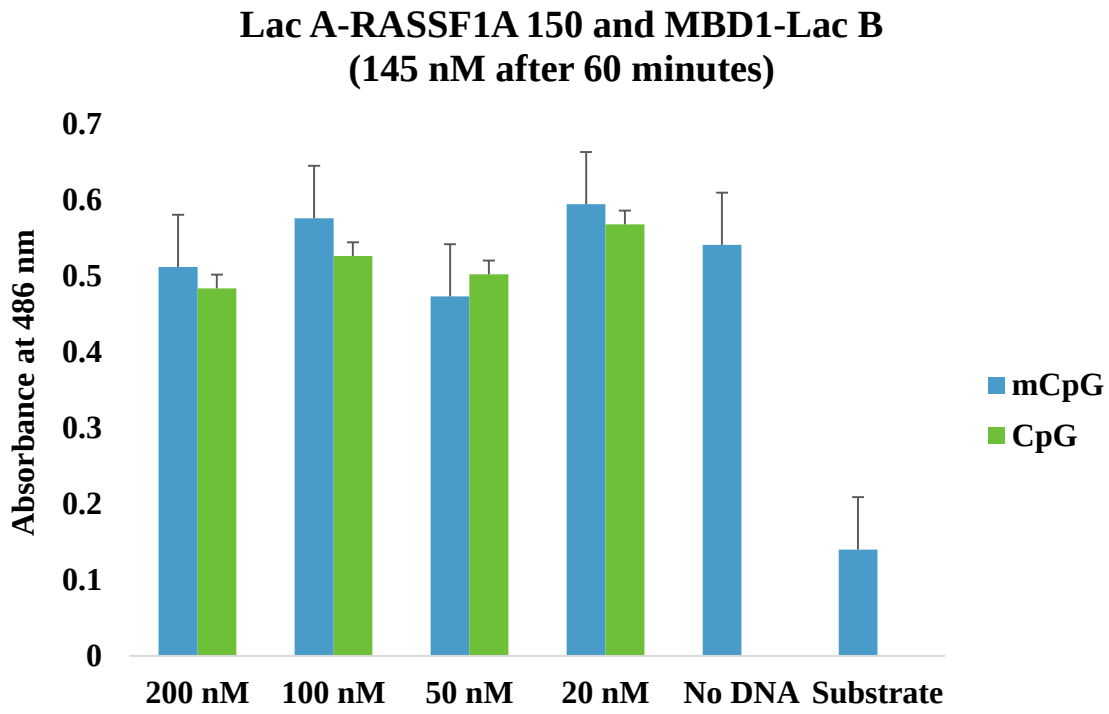

Figure 12: Nitrocefin assay using Lac A-RASSF1A 150 and MBD1-Lac B of concentration 145 nM with DNA concentration 200, 100, 50, 20 nM.

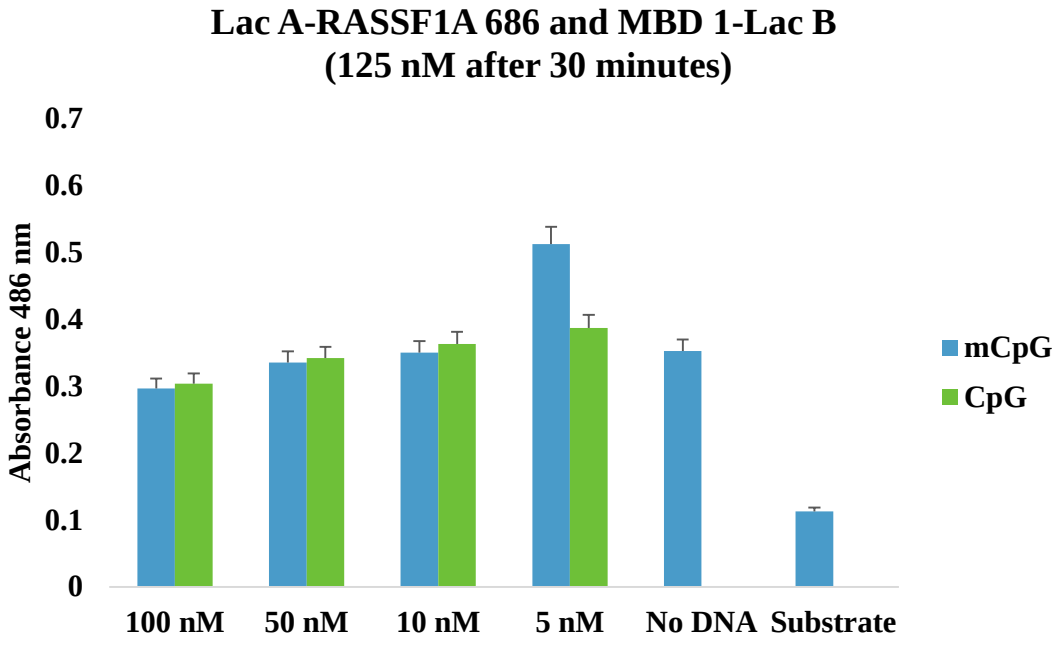

Figure 13: Nitrocefin assay using Lac A-RASSF1A 686 and MBD1-Lac B of concentration 125 nM with DNA concentration 100, 50, 10, 5 nM.

From the above sets of solution based nitrocefin assays, following observations were made as per each protein:

Lac A-RASSF1A 150

This ZFP did not show any qualitative signals for binding in the solution based assay. It was observed that increasing DNA concentration also did not improve in the signals and increasing protein concentration was also not very useful. It is unclear why higher signals were not observed by increasing protein concentrations.

Lac A-RASSF1A 686

Although this construct gave the highest signals for non-specific binding, the assay revealed some details of MBD's specificity by showing lower signals for 5 nM CpG and higher for mCpG. Lowering DNA and protein concentration did not have any affect on lowering the signals. Oligonucleotides having the mCpG group showed an unexpected lowering in signals. We could speculate that hairpin DNA oligonucleotides showed the lowest signals because their structure pushed the target and mCpG site far from each other, making it difficult for the split Lac domains to reassemble. It was predicted that mCpG should be positioned on the sense strand of the DNA while ZF target should be positioned on the anti-sense strand because when both mCpG and ZFP target site are on the sense strand, coiling of the DNA might push the target side away and on the other side of the DNA, causing Lac A and Lac B to be distant. Target sequence on the anti-sense strand could ensure that the target falls on the same side as the mCpG group, making it easier for the split Lac domains to reassemble. It is necessary that both mCpG and target site fall in the major groove of the DNA. The proteins are relatively larger in size and need more space to get accommodated. The major groove provides ample space for split Lac domain reassembly and protein-DNA binding. Our data suggests that hairpin DNA structures

also have a tendency to increase the distance between mCpG and target site. In some cases, non-methylated oligonucleotides also show signals due to self-reassembly of Lac A and Lac B.

Our results suggest inefficiency of the solution based assay to be caused by background signals due to the self-reassembly of Lac A and Lac B at high protein concentrations. Therefore, we implemented another method by making use of a protein array on the surface, which would allow for removal of unbound and weakly bound molecules.

5. Experiments from zinc finger protein array

Table 5: Set of experiments from the ZFP array

Assay #	Lac A-ZFP protein 2.5 μM	MBD-Lac B protein 2.5 μM	DNA concentration (nM)	Oligonucleotide sequence #
1.	RASSF1A 150	MBD1	250, 50, 10	149 mCpG 2bp-3 149 CpG 2bp-3
2.	RASSF1A 150	MBD1	250, 10	149 mCpG 3bp-3 149 mCpG 3bp-3 149 mCpG 2bp-3 149 CpG 2bp-3
3. a	RASSF1A 150	MBD1 MBD2	250, 50, 10	149 mCpG 2bp-3 149 CpG 2bp-3
b	RASSF1A 150	MBD 1 MBD 2	250, 50 ,10	149 mCpG 3bp-3 149 mCpG 3bp-3
4.	RASSF1A 407	MBD1 MBD2	50,10	402 mCpG- 4 402 CpG- 4

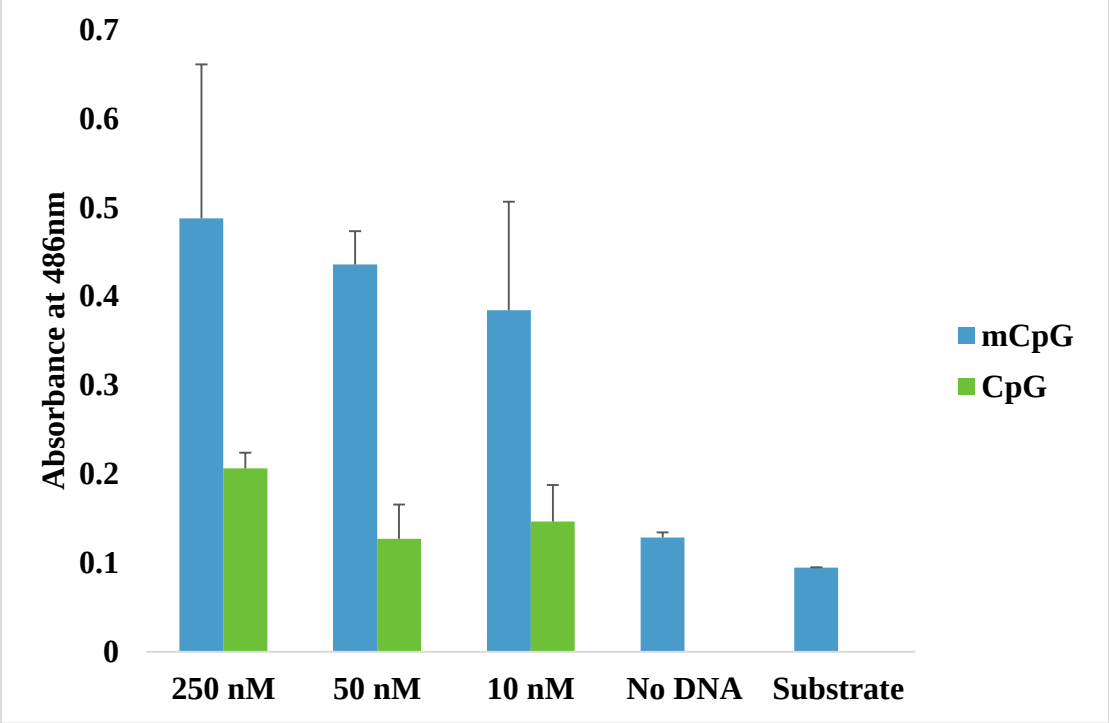

Figure 14: Nitrocefin assay using Lac A-RASSF1A 150 and MBD 1-Lac B at protein concentration of 2.5 μM. Determination of DNA binding at different DNA concentrations. (Assay 1)

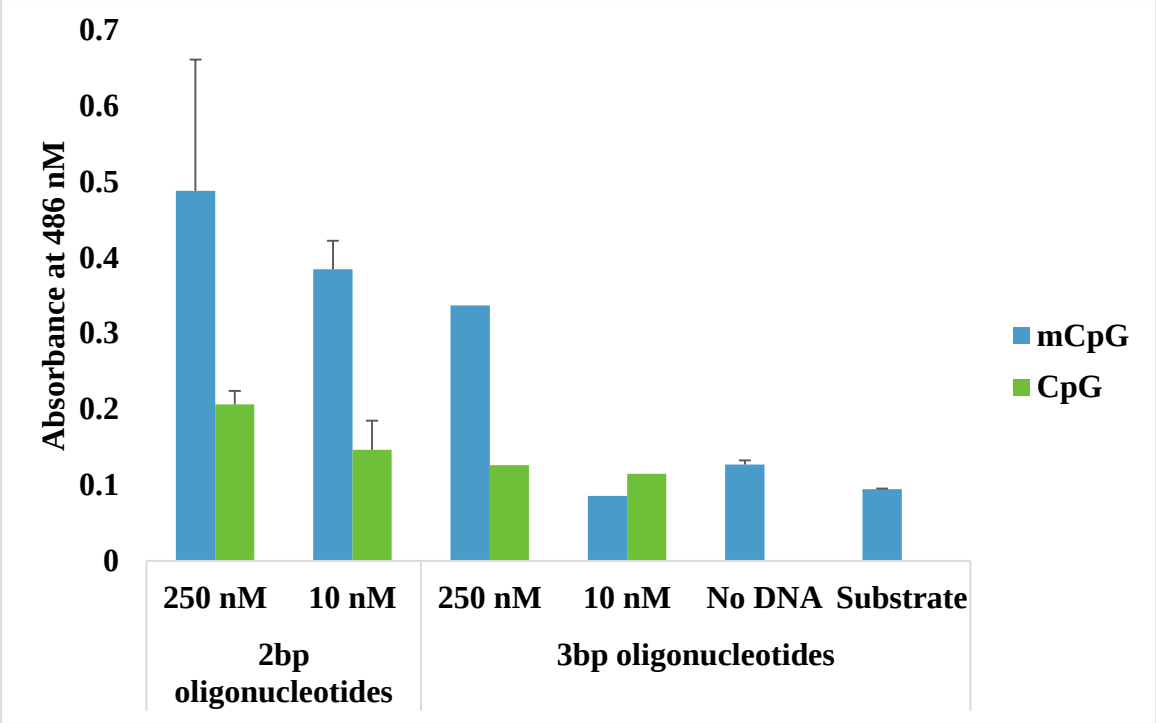

Figure 15: Nitrocefin assay for signal comparison of 2 bp and 3 bp spacing oligonucleotides with concentrations of 250 and 10 nM, using Lac A-RASSF1A 150 and MBD1-Lac B. (Assay 2)

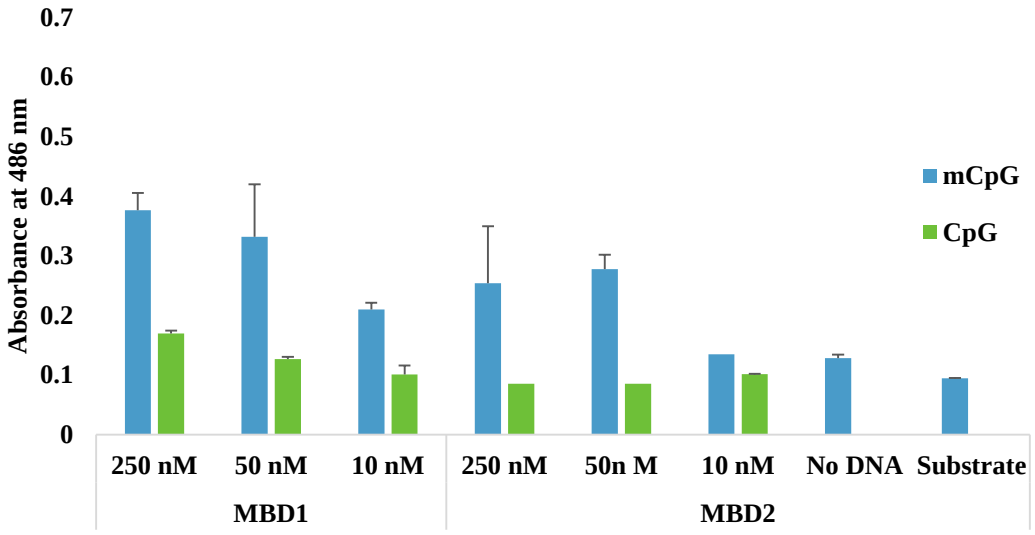

Figure 16: Nitrocefin assay show comparison of MBD 1 and MBD 2 with Lac A-RASSF1A 150 with oligonucleotides of 2 bp spacing. Proteins are of the concentration 250 nM. (Assay 3a)

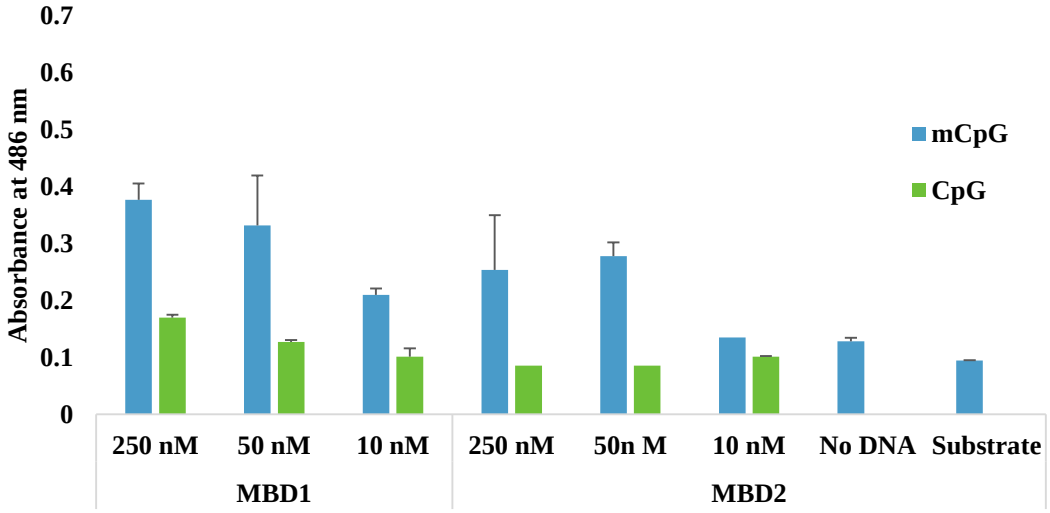

Figure 17: Nitrocefin assay showing comparison of MBD 1 and MBD 2 with Lac A-RASSF1A 150 with oligonucleotides of 3 bp spacing. Proteins are of the concentration 250 nM. (Assay 3b)

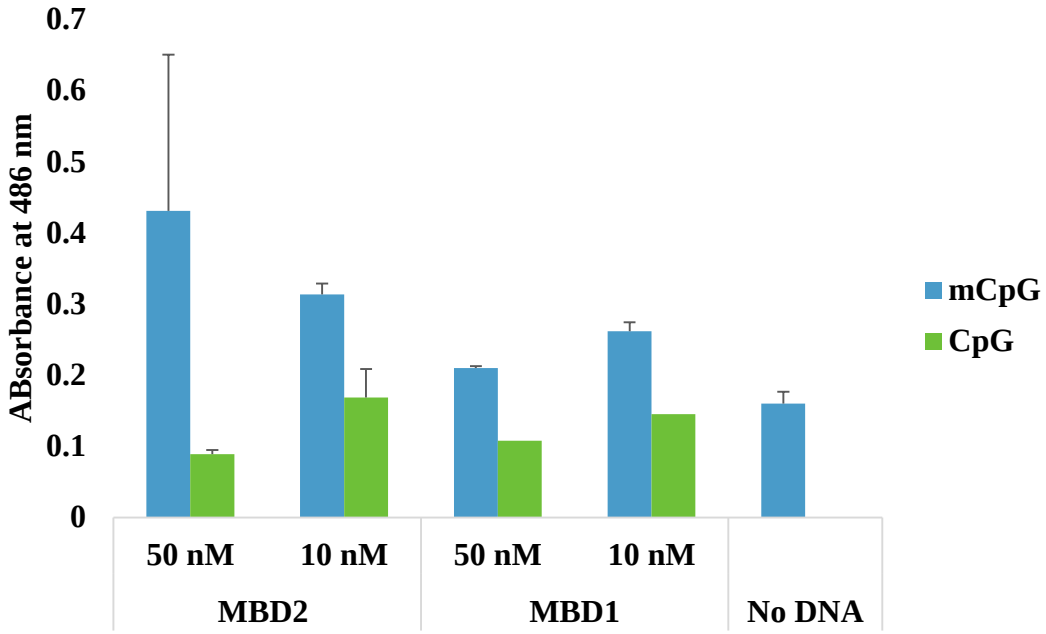

Figure 18: Nitrocefin assay showing comparison of MBD 1 and MBD 2 with Lac A-RASSF1A 407 and DNA concentration of 50 and 10 nM. (Assay 4)

Assay 1

Lac A- RASSF1A 150 in figure 15, shows dose dependent signal with MBD1-Lac B at protein concentration 2.5 μM and oligonucleotide concentrations of 250, 50, and 10 nM. This suggests that ZFP RASSF1A 150 has high selectivity for the system. MBD1 being a strong repressor of the methylated DNA, binds tightly and specifically to the methylated target. The position of mCG and ZFP target is most suitable in this case, showing great signal difference between methylated and non-methylated DNA. Low signals for substrate alone and DNA suggest low background.

Assay 2

Two oligonucleotides with different spacing between mCpG and ZF target were used. One was positioned 2 bp apart and the other was 3 bp apart. The results from figure 16 suggested that 2 bp shows higher efficiency than 3 bp, probably owing to the fact that it is situated closer to the axis of the DNA double helix resulting in more space to allow effective reassembly of Lac A and Lac B. Even though the difference in space for 2 bp and 3 bp is not as much, a difference in 1 bp near the vertices of the double helix may result in a relatively big difference in the position of the ZFP target. This may hinder the spontaneous association of Lac A and Lac B due to their geometric conformation and location. DNA concentrations of 250 nM and 10 nM show accurate difference in signal between methylated DNA and non-methylated DNA as well as lower signals for the 3 bp oligonucleotide DNA.

Assay 3a, 3b, and 4

From figures 17 and 18, the two different oligonucleotides mentioned above were checked for signals using both MBD1 and MBD2 proteins. In the case of RASSF1A 150, in both instances, MBD1 showed more effective binding results in the mCpG-SEER-Lac system. However, from figure 18 we observed that MBD2 shows higher signals than MBD1 in the mCpG SEER Lac system with RASSF1A 407.

MBD1 and MBD2 are the only methyl binding domains that respond to methylated DNA (51). In our system, both MBD1 and MBD2 were capable of differentiating methylated target from non-methylated target.

Effects of oligonucleotide structure on the assay

The oligonucleotides first designed were 31 bp long and had both the methyl cytosine and target on the sense strand. It can be inferred that when the target is on the antisense strand, the helix formation would cause the target and the methyl cytosine (mC) to fall in the major groove. If the opposite was the case, then the target and mC would be farther away from each other and one of them would fall in the minor groove. This distancing would make it impossible for the Lac A and Lac B to reassemble upon binding of MBD and ZFP. The length of the oligonucleotide may change the origin of the helix and cause the mC to fall in the minor groove, possibly preventing MBD from binding to mCG. Since arginine in MBD makes contacts with G via two hydrogen bonds and hydrophobic interaction with G, it is a primary requirement that mC should be accompanied with a G (43). Lastly in figure 18, we designed a 39 bp oligonucleotide, with the first 12 bp of random sequence followed by mC. After leaving a 2-3 bp spacing, the target was placed on the anti-sense strand. Starting with a 12 bp random sequence ensures completion of one turn of the DNA double helix. We could hence assume that the mC and target site fall in the major groove. Our data also suggests that 2 bp spacing provides favorable space and reduction in distance between Lac A and Lac B, so they reassemble more efficiently than 3 bp spacing.

Other factors that could influence successful discrimination between methylated and non-methylated oligonucleotide are the binding incubation time and washing steps involved in the ZFP array. Immobilization of the ZFP on the cyclic olefin copolymer (COC) chip via hydrophobic interactions confirms accessibility of ZFPs. Incubation time for DNA binding allows protein-DNA interaction to take place. ZFPs binds to its target site and unbound DNA washes off in the washing step. This

gives a clear stage for MBD to bind to the DNA which is already bound to the ZFP on its target site. Simultaneous binding of MBD and ZFP to DNA leads to reassembly of Lac A and Lac B. ZFP array allows binding of the ZFP and MBD to cognate DNA. In case of a non-methylated DNA, the ZFP array will allow binding of ZFP to target but will wash away the unbound MBD, resulting in relatively low signal.

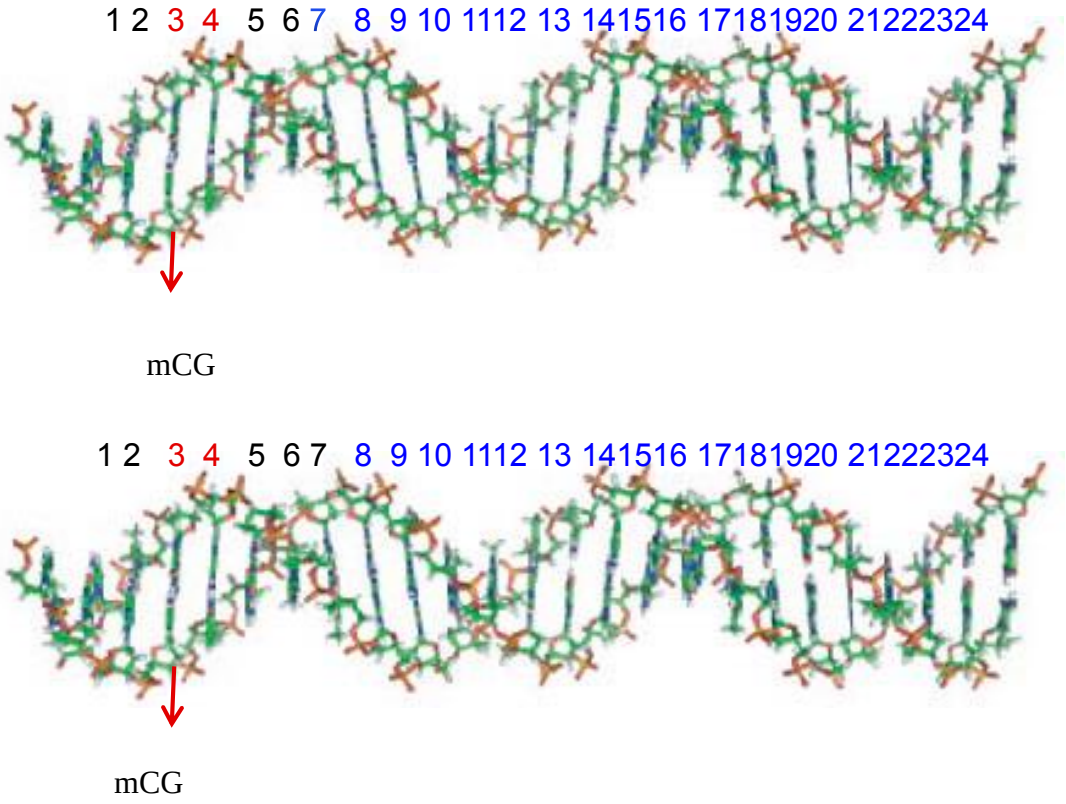

Figure 19: Oligonucleotide sequence highlighting methylated CG dinucleotide (mCG) (red) and ZFP target (blue).

CHAPTER 4

CONCLUSION

In conclusion, mCpG SEER-Lac system is an efficient system to detect a desired methylated DNA. The idea of using MBD proteins serves as an excellent detector system that binds to the mCpG on the DNA (51). However, the solution based assays from figures 12 and 13, did not show any signal difference between methylated and non-methylated DNA. It also did not reveal any dose dependent signals. This encouraged us to implement the use of a hydrophobic COC chip based nitrocefin assay. Immobilization of the ZFP and sufficient incubation time for DNA-ZFP binding allows elimination of unbound DNA and availability of ZFP-DNA complex. This makes it easy for the MBD to find the methylated DNA and bind to it. ZFP array shows a great difference in signal between methylated and non-methylated DNA. Figure 14, 15 and 17 show a dose dependent signal for methylated target. Signals for reactions without target DNA were also reduced. This diminishes the probability of false positives. These effects can be explained on the basis of the geometry attained by the DNA after binding with the ZFP and position of the target and methyl. The key idea was to make enough space in the major groove of the DNA to allow binding of both MBD on the mCpG and ZFP on its target. Simultaneous binding allows Lac A and Lac B to be positioned in close proximity facilitating activation on reassembly of the two inactive fragments. This assumption was corroborated by the results from assay 2 in figure 15, where, the ZF target was placed with a 2 bp spacing next to the mCG in the oligonucleotides. Results from the 3 bp oligonucleotide could be lower because of increase in the distance between mCpG and ZF target. In the 39 bp oligonucleotide described earlier, it is possible that the

target gets positioned at the center of the loop (or towards the end of the major groove) making it difficult for the LacA and Lac B to reassemble.

MBD1 and MBD2 both show binding to mammalian mCpG (47). However, many other experiments also suggest that MBD1 and MBD2 show higher preference for mCpG than other methyl binding proteins (51). This was the reason for selection of MBD 1 and MBD2 proteins in our assay. Our results from figures 16 and 17 suggest that both MBD1 and MBD2 are specific in the mCpG SEER-Lac system in binding to its methylated target sequence. Since both ZFPs have 3 GNN targeting ZF modules in their structures, both share a distributed sense of binding across the cognate DNA sequence. EMSA results from figure 11 also suggest that both RASSF1A 150 and 407 show good binding affinity towards their target confirming the intensity of signals obtained. Large signal difference in the methylated and non-methylated DNA show that the system is effective in detection of methylated DNA and reduces the risk of attaining false positives. Rectification of this error makes mCpG- SEER Lac a better detection system than other methylation detection system like Bisulfite sequencing.

CHAPTER 5

FUTURE DIRECTION

Detection of cancer at the early stage is very important to help in prognosis. Cancer biomarkers are the most suitable target for detection of cancer. In case of bladder cancer, a panel with RASSF1A, APC and p14 serves as a reliable set of biomarkers to detect bladder cancer in urine (29,52). Efficiency of such panels with combination of different biomarkers have been studied using other methods (29). Our approach enables us to detect specific DNA sequences in the RASSF1A for bladder cancer. A point of-care device could be developed with custom designed ZFP targeting specific regions in multiple genes of a cancer biomarker. Other biomarkers related to the same cancer can also be tested alongside. This panel for detection of different cancer biomarkers may provide information needed to indicate the presence of a tumor. There are biomarkers which can be detected in the body like fluid blood or urine from concerned tissue (27). This diagnostic test is more beneficial over a penetrative procedure to obtain biopsy material for the detection of cancer.

REFERENCES

1. Kazanets, A., Shorstova, T., Hilmi, K., Marques, M. and Witcher, M. (2016) Epigenetic silencing of tumor suppressor genes: Paradigms, puzzles, and potential. Biochim Biophys Acta, **1865**, 275-288.
2. Wang, T., Liu, H., Chen, Y., Liu, W., Yu, J. and Wu, G. (2009) Methylation associated inactivation of RASSF1A and its synergistic effect with activated K-Ras in nasopharyngeal carcinoma. J Exp Clin Cancer Res, **28**, 160.
3. Rajalingam, K., Schreck, R., Rapp, U.R. and Albert, S. (2007) Ras oncogenes and their downstream targets. Biochim Biophys Acta, **1773**, 1177-1195.
4. Eckfeld, K., Hesson L Fau - Vos, M.D., Vos Md Fau - Bieche, I., Bieche I Fau - Latif, F., Latif F Fau - Clark, G.J. and Clark, G.J. (2004) RASSF4/AD037 is a potential ras effector/tumor suppressor of the RASSF family. Cancer, **64**, 8688-8693.
5. Donninger, H., Vos, M.D. and Clark, G.J. (2007) The RASSF1A tumor suppressor. J Cell Sci, **120**, 3163-3172.
6. Akino, K., Toyota M Fau - Suzuki, H., Suzuki H Fau - Mita, H., Mita H Fau - Sasaki, Y., Sasaki Y Fau - Ohe-Toyota, M., Ohe-Toyota M Fau - Issa, J.-P.J., Issa Jp Fau - Hinoda, Y., Hinoda Y Fau - Imai, K., Imai K Fau - Tokino, T. and Tokino, T. (2005) The Ras effector RASSF2 is a novel tumor-suppressor gene in human colorectal cancer. Gastroenterology, **129**, 156-169.
7. Tommasi, S., Dammann R Fau - Jin, S.-G., Jin Sg Fau - Zhang, X.-f., Zhang Xf Fau - Avruch, J., Avruch J Fau - Pfeifer, G.P. and Pfeifer, G.P. (2002) RASSF3 and NORE1: identification and cloning of two human homologues of the putative tumor suppressor gene RASSF1. Oncogene, **21**, 2713-2720.
8. Vavvas, D., Li X Fau - Avruch, J., Avruch J Fau - Zhang, X.F. and Zhang, X.F. (1998) Identification of Nore1 as a potential Ras effector. J. Biol. Chem. , **273**, 5439-5442.
9. Khokhlatchev, A., Rabizadeh S Fau - Xavier, R., Xavier R Fau - Nedwidek, M., Nedwidek M Fau - Chen, T., Chen T Fau - Zhang, X.-f., Zhang Xf Fau - Seed, B., Seed B Fau - Avruch, J. and Avruch, J. (2002) Identification of a novel Ras-regulated proapoptotic pathway. Curr. Biol. , **12**, 253-265.
10. Praskova, M., Khoklatchev A Fau - Ortiz-Vega, S., Ortiz-Vega S Fau - Avruch, J. and Avruch, J. (2004) Regulation of the MST1 kinase by autophosphorylation, by the growth inhibitory proteins, RASSF1 and NORE1, and by Ras. Biochem. J. , **381**, 353-362.
11. Katagiri, K., Maeda A Fau - Shimonaka, M., Shimonaka M Fau - Kinashi, T. and Kinashi, T. (2003) RAPL, a Rap1-binding molecule that mediates Rap1-induced adhesion through spatial regulation of LFA-1. nat. Immunol. , **4**, 741-748.
12. Allen, N.P., Donninger H Fau - Vos, M.D., Vos Md Fau - Eckfeld, K., Eckfeld K Fau - Hesson, L., Hesson L Fau - Gordon, L., Gordon L Fau - Birrer, M.J., Birrer Mj Fau - Latif, F., Latif F Fau - Clark, G.J. and Clark, G.J. (2007) RASSF6 is a novel member of the RASSF family of tumor suppressors. Oncogene.
13. Hull, J., Rowlands, K., Lockhart, E., Sharland, M., Moore, C., Hanchard, N. and Kwiatkowski, D.P. (2004) Haplotype mapping of the bronchiolitis susceptibility locus near IL8. Hum Genet, **114**, 272-279.
14. Meng, W., Huebner, A., Shabsigh, A., Chakravarti, A. and Lautenschlaeger, T. (2012) Combined RASSF1A and RASSF2A Promoter Methylation

Analysis as Diagnostic Biomarker for Bladder Cancer. Molecular Biology International, **2012**, 8.
15. Amin, K.S. and Banerjee, P.P. (2012) The cellular functions of RASSF1A and its inactivation in prostate cancer. Journal of Carcinogenesis, **11**, 3.
16. Agathanggelou, A., Cooper Wn Fau - Latif, F. and Latif, F. (2005) Role of the Ras-association domain family 1 tumor suppressor gene in human cancers. Cancer Research, **65**.
17. Dammann, R., Schagdarsurengin U Fau - Seidel, C., Seidel C Fau - Strunnikova, M., Strunnikova M Fau - Rastetter, M., Rastetter M Fau - Baier, K., Baier K Fau - Pfeifer, G.P. and Pfeifer, G.P. The tumor suppressor RASSF1A in human carcinogenesis: an update.
18. Rappaport, R. (1996) Cytokinesis in Animal Cells. Cambridge University Press.
19. Ura, S., Nishina, H., Gotoh, Y. and Katada, T. (2007) Activation of the c-Jun N-terminal kinase pathway by MST1 is essential and sufficient for the induction of chromatin condensation during apoptosis. Mol Cell Biol, **27**, 5514-5522.
20. Shivakumar, L., Minna, J., Sakamaki, T., Pestell, R. and White, M.A. (2002) The RASSF1A tumor suppressor blocks cell cycle progression and inhibits cyclin D1 accumulation. Mol Cell Biol, **22**, 4309-4318.
21. Pfeifer, G.P. and Dammann, R. (2005) Methylation of the tumor suppressor gene RASSF1A in human tumors. Biochemistry **70**, 576-583.
22. Foley, C.J., Freedman, H., Choo, S.L., Onyskiw, C., Fu, N.Y., Yu, V.C., Tuszynski, J., Pratt, J.C. and Baksh, S. (2008) Dynamics of RASSF1A/MOAP-1 association with death receptors. Mol Cell Biol, **28**, 4520-4535.
23. Hendrich, B. and Tweedie, S. (2003) The methyl-CpG binding domain and the evolving role of DNA methylation in animals. Trends in Genetics **19**.
24. Zangi, R., Arrieta, A. and Cossio, F.P. (2010) Mechanism of DNA methylation: the double role of DNA as a substrate and as a cofactor. J Mol Biol, **400**, 632-644.
25. Ohki, I., Shimotake, N., Fujita, N., Jee, J.-G., Ikegami, T., Nakao, M. and Shirakawa, M. (2001) Solution Structure of the Methyl-CpG Binding Domain of Human MBD1 in Complex with Methylated DNA. Cell, **105**, 487-497.
26. Fuks, F., Burgers, W.A., Brehm, A., Hughes-Davies, L. and Kouzarides, T. (2000) DNA methyltransferase Dnmt1 associates with histone deacetylase activity. Nat Genet, **24**, 88-91.
27. Hesson, L.B., Cooper, W.N. and Latif, F. (2007) The role of RASSF1A methylation in cancer. Dis Markers, **23**, 73-87.
28. Loginov, V.I., Maliukova, A.V., Seregin Iu, A., Khodyrev, D.S., Kazubskaia, T.P., Ermilova, V.D., Gar'kavtseva, R.F., Kiselev, L.L., Zabarovskii, E.R. and Braga, E.A. (2004) [Methylation of the promoter region of the RASSF1A gene, a candidate tumor suppressor, in primary epithelial tumors]. Mol Biol (Mosk), **38**, 654-667.
29. Kandimalla, R., van Tilborg, A.A. and Zwarthoff, E.C. (2013) DNA methylation-based biomarkers in bladder cancer. Nat Rev Urol, **10**, 327-335.
30. Dammann, R., Schagdarsurengin, U., Seidel, C., Strunnikova, M., Rastetter, M., Baier, K. and Pfeifer, G.P. (2005) The tumor suppressor RASSF1A in human carcinogenesis: an update. Histol Histopathol, **20**, 645-663.

31. Leon, S.A., Shapiro, B., Sklaroff, D.M. and Yaros, M.J. (1977) Free DNA in the serum of cancer patients and the effect of therapy. Cancer Res, **37**, 646-650.
32. Tian, Y., Hou, Y., Zhou, X., Cheng, H. and Zhou, R. (2011) Tumor suppressor RASSF1A promoter: p53 binding and methylation. PLoS One, **6**, e17017.
33. Seeman, N.C., Rosenberg, J.M. and Rich, A. (1976) Sequence-specific recognition of double helical nucleic acids by proteins. Proceedings of the National Academy of Sciences of the United States of America, **73**, 804-808.
34. Klug, A. (2010) The discovery of zinc fingers and their applications in gene regulation and genome manipulation. Annu Rev Biochem, **79**, 213-231.
35. Lee, M.S., Gippert, G.P., Soman, K.V., Case, D.A. and Wright, P.E. (1989) Three-dimensional solution structure of a single zinc finger DNA-binding domain. Science, **245**, 635-637.
36. Kim, M.S., Stybayeva, G., Lee, J.Y., Revzin, A. and Segal, D.J. (2011) A zinc finger protein array for the visual detection of specific DNA sequences for diagnostic applications. Nucleic Acids Res, **39**, e29.
37. Klug, A. (2010) The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation. Q Rev Biophys, **43**, 1-21.
38. Park, K.S., Lee, D.K., Lee, H., Lee, Y., Jang, Y.S., Kim, Y.H., Yang, H.Y., Lee, S.I., Seol, W. and Kim, J.S. (2003) Phenotypic alteration of eukaryotic cells using randomized libraries of artificial transcription factors. Nat Biotechnol, **21**, 1208-1214.
39. Segal, D.J., Beerli, R.R., Blancafort, P., Dreier, B., Effertz, K., Huber, A., Koksch, B., Lund, C.V., Magnenat, L., Valente, D. et al. (2003) Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins. Biochemistry, **42**, 2137-2148.
40. Bhakta, M.S. and Segal, D.J. (2010) The generation of zinc finger proteins by modular assembly. Methods in molecular biology, **649**, 3-30.
41. Carroll, D., Morton, J.J., Beumer, K.J. and Segal, D.J. (2006) Design, construction and in vitro testing of zinc finger nucleases. Nature protocols, **1**, 1329-1341.
42. Wright, D.A., Thibodeau-Beganny, S., Sander, J.D., Winfrey, R.J., Hirsh, A.S., Eichtinger, M., Fu, F., Porteus, M.H., Dobbs, D., Voytas, D.F. et al. (2006) Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. Nature protocols, **1**, 1637-1652.
43. Zou, X., Ma, W., Solov'yov, I.A., Chipot, C. and Schulten, K. (2012) Recognition of methylated DNA through methyl-CpG binding domain proteins. Nucleic Acids Res, **40**, 2747-2758.
44. Ohki, I., Shimotake, N., Fujita, N., Jee, J., Ikegami, T., Nakao, M. and Shirakawa, M. (2001) Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. Cell, **105**, 487-497.
45. Meehan, R.R., Lewis Jd Fau - McKay, S., McKay S Fau - Kleiner, E.L., Kleiner El Fau - Bird, A.P. and Bird, A.P. (1989) Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell, **58**, 499-507.
46. Lewis, J.D., Meehan Rr Fau - Henzel, W.J., Henzel Wj Fau - Maurer-Fogy, I., Maurer-Fogy I Fau - Jeppesen, P., Jeppesen P Fau - Klein, F., Klein F Fau - Bird, A. and Bird, A. (1992) Purification, sequence, and cellular localization

of a novel chromosomal protein that binds to methylated DNA. Cell, **69**, 905-914.
47. Hendrich, B. and Bird, A. (1998) Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol Cell Biol, **18**, 6538-6547.
48. Chow, L.S., Lam, C.W., Chan, S.Y., Tsao, S.W., To, K.F., Tong, S.F., Hung, W.K., Dammann, R., Huang, D.P. and Lo, K.W. (2006) Identification of RASSF1A modulated genes in nasopharyngeal carcinoma. Oncogene, **25**, 310-316.
49. Wong, E., Yang, K., Kuraguchi, M., Werling, U., Avdievich, E., Fan, K., Fazzari, M., Jin, B., Brown, A.M., Lipkin, M. et al. (2002) Mbd4 inactivation increases Cright-arrowT transition mutations and promotes gastrointestinal tumor formation. Proc Natl Acad Sci U S A, **99**, 14937-14942.
50. Ooi, A.T., Stains, C.I., Ghosh, I. and Segal, D.J. (2006) Sequence-enabled reassembly of beta-lactamase (SEER-LAC): a sensitive method for the detection of double-stranded DNA. Biochemistry, **45**, 3620-3625.
51. Badran, A.H., Furman, J.L., Ma, A.S., Comi, T.J., Porter, J.R. and Ghosh, I. (2011) Evaluating the global CpG methylation status of native DNA utilizing a bipartite split-luciferase sensor. Anal Chem, **83**, 7151-7157.
52. Hoque, M.O., Begum, S., Topaloglu, O., Chatterjee, A., Rosenbaum, E., Van Criekinge, W., Westra, W.H., Schoenberg, M., Zahurak, M., Goodman, S.N. et al. (2006) Quantitation of promoter methylation of multiple genes in urine DNA and bladder cancer detection. J Natl Cancer Inst, **98**, 996-1004.

ABBREVIATIONS

1. **ZFP- Zinc Finger Protein**
2. **TSG- Tumor suppressor gene**
3. **RASSF1A- Ras- associated family 1 subfamily A**
4. **Nore1a- RASSF1A homolog**
5. **SARAH- sav/RASSF/Hpo domain**
6. **CpG- Cytosine Guanine dinucleotide**
7. **mCpG- methylated cytosine guanine dinucleotide**
8. **GTPase- guanosine triphosphatase**
9. **PMCa- plasma membrane calcium ATPase**
10. **MOAP-1- Modulator of apoptosis 1**
11. **SAV- Salvador protein**
12. **LATs ½- Linker for activation of T-cells**
13. **MST1- Macrophage stimulating 1 protein**
14. **CNK1- Connector enhancer of kinase suppressor of RAS 1**
15. **BAX- BCL2- Associated X protein**
16. **MAP18- Microtubule associated protein**
17. **APC- Adenomatous polyposis coli**
18. **JNK-** c-Jun-NH2-Kinase
19. **TNFR- Tumor necrosis factor receptor**
20. **DNMT- DNA methyltransferase**
21. **MBD- Methyl binding domain**
22. **Val- valine**
23. **Arg- Arginine**
24. **Tyr- Tyrosine**

25. Ser- Serine

26. NMR- Nuclear magnetic resonance

27. Phe- Phenyl alanine

28. Ala- Alanine

29. DBD- Dna Binding domain

30. SEER- Sequence Enabled reassembly

31. IPTG- Isopropyl β-D-1-thiogalactopyranoside

32. ZBA- Zinc buffer A

33. TCEP- tris(2-carboxyethyl) phosphine

34. DTT- dithiothreitol

35. NaCl- Sodium chloride

36. BSA- Bovine serum albumin

37. COC- Cyclic olefin copolymer

38. nM- nanomolar

39. nm- nonometers

40. μl- microliter

41. μM- micromolar

42. °C- Degree centigrade

43. bp- basepair