MQP-BIO-DSA-9389

Analysis of Matrix Attachment Region in an Experimental Vector, pA205

A Major Qualifying Project Report

Submitted to the Faculty of the

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the

Degree of Bachelor of Science

By

Alexia T. Kedves

March 14, 2005

APPROVED:

Wendy Gion

Biologics Department,
Cell Line Development Lab
Abbott Bioresearch Center
Major Advisor

David S. Adams, Ph.D.

Professor, Biology and Biotech
WPI Project Advisor

ABSTRACT

Recent literature has shown that Matrix and Scaffold Attachment Regions (MARs or SARs) are DNA domains that enhance transcriptional frequency and increase gene expression levels in cells. Research has shown that cell lines containing MARs or SARs are more stable and more productive. Researchers at Abbott Bioresearch Center in Worcester, MA have found two possible MAR regions in an antibody producing cell line. In order to determine whether these regions function as MARs, and if the observed high levels of expression are specific to this cell line, the sequences in question were inserted upstream of the reporter gene, Enhanced Green Fluorescent Protein (EGFP) and tested using FACS analysis. The data indicated the sequences under investigation indeed increased expression of EGFP in stable cell lines, indicating that these sequences can be used to improve transfection and establish high expressing, stable CHO cell lines.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to thank all of the people who have helped me complete this project at Abbott Bioresearch Center. First, I would like to thank Gerald Carson, head of the Cell Line Development Group, for his insight and guidance. Next, I would like to thank Wendy Gion for all of her guidance throughout the project, being my main advisor, and for all her help in the laboratory. Along with Wendy and Jerry, I would like to thank Asma Ahmed, Shaona Fang, and Dr. Alison Mastrangelo for help in the lab, and Carrie Goodreau for supplying me with the hCMV-GFP positive control plasmid. Also, I wish to thank Dean Regier for sequencing the parent and genomic constructed vectors, and Sukumari Mohan in the numerous FACS analyses performed throughout the project. Finally, I would like to thank Dr. David Adams, professor and advisor of this Major Qualifying Project at Worcester Polytechnic Institute, for helping with the initiation and completion of the project.

BACKGROUND

Nucleosome Remodeling in Chromatin

Chromatin is comprised of genomic DNA wrapped into nucleosomes which compact to form a condensed chromatin fiber inside the nucleus (Figure 1). Structural changes within chromatin have been found to play a central role in the control of gene expression.

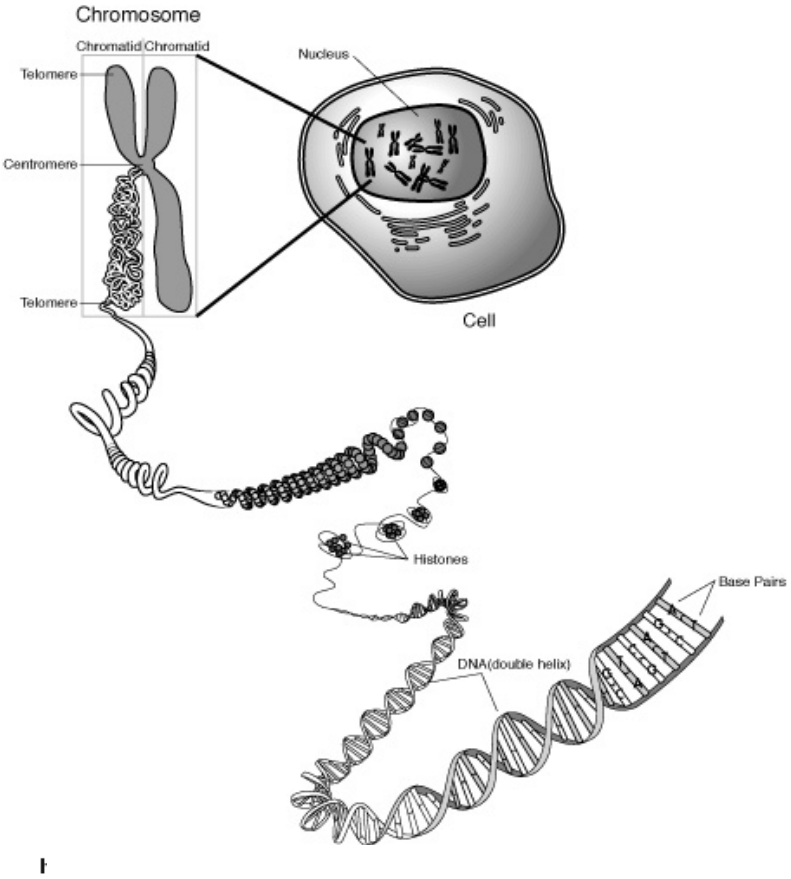

Figure 1- Structural Organization of Chromatin- DNA wraps around the histone "beads" to form nucleosomes which, when compacted, form chromatin (Access Excellence, 2005)

Two types of enzymes with the ability to change chromatin structure have been identified as the factors changing chromatin structure: 1) ATP hydrolyzing enzymes that remodel chromatin at the nucleosome level, and 2) enzymes with the abiltiy to modify specific amino acid residues, primarily on the nucleosomal histone tails (Kouzarides, 2002). It is

important to note that both of these activities function by acting on the nucleosomes. The building blocks of chromatin, nucleosomes, are the detailed structure within which eukaryotic cells organize their genome, while still allowing accessibility to regulatory factors. Each nucleosome consists of ~150 base pairs of DNA wrapped in spools around the histone octamers. They are connected to other nucleosomes by ~50 base pair stretches of "linker" DNA (Becker and Hörz, 2002). Every histone has an N-terminal tail, that is open to post translational modifications including acetylation, phosphorylation and methylation, all of which affect gene regulation (Becker and Hörz, 2002). These modifications will be described more in depth.

Becker and Hörz define nucleosome remodeling as the process with which histone-DNA interactions are altered inside a nucleosome in an ATP-dependent manner, usually rendering nucleosomal DNA more accessible for protein interaction. Packaging of promoter DNA into the nucleosome structure has been found to inhibit transcription in vitro. A change in nucleosome structure or internucleosomal associations reverses this inhibition (Boeger et. al, 2003) and causes transcriptional activation. This means that nucleosomal DNA can be made accessible by unfolding the nucleosome.

ATP Dependent Nucleosome Remodeling Factors

ATP dependent nucleosome remodeling factors enable the nucleosomes to become more accessible to protein interactions (Becker and Hörz, 2002). The following figure (Figure 2) shows the organization of the histones and DNA into nucleosomes.

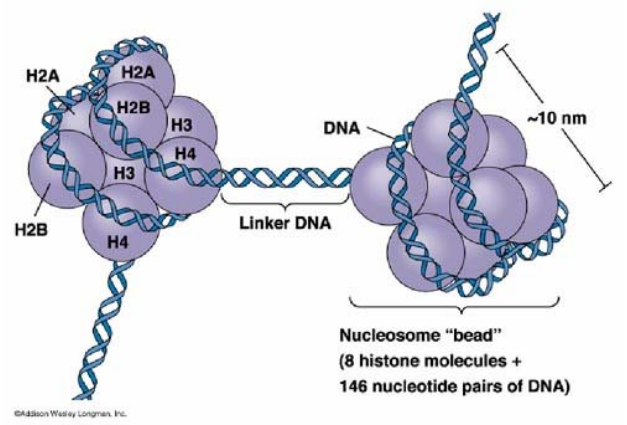

Figure 2- Nucleosome Structure- 146 bp of DNA associates with a histone octamer to form one nucleosome. In between each nucleosome is about 50 bp of linker DNA. Histone H1 is thought to bind to linker DNA, functioning in nucleosome compaction. (Addison Wesley Longman, 1999)

Nucleosome remodeling factors function in activation and repression of transcription during the cell cycle and cell differentiation, and therefore they function in the development of multicellular organisms. Remodeling is caused by the relocation of histone octamers. Within the Swi/Snf family, all remodeling factors bear the ATPase subunit of the Swi2/Snf2 family (Becker and Hörz, 2002). The enzyme subfamilies of Swi/Snf2 are explained briefly below, with differences dependent on their features other than the ATPase domain. Primary focus will be on Swi/Snf complexes, ISWI/SNF2L containing factors, and the CHD complexes because little is known yet about the others. Figure 3 organizes the families of ATPases and their subsequent subfamilies.

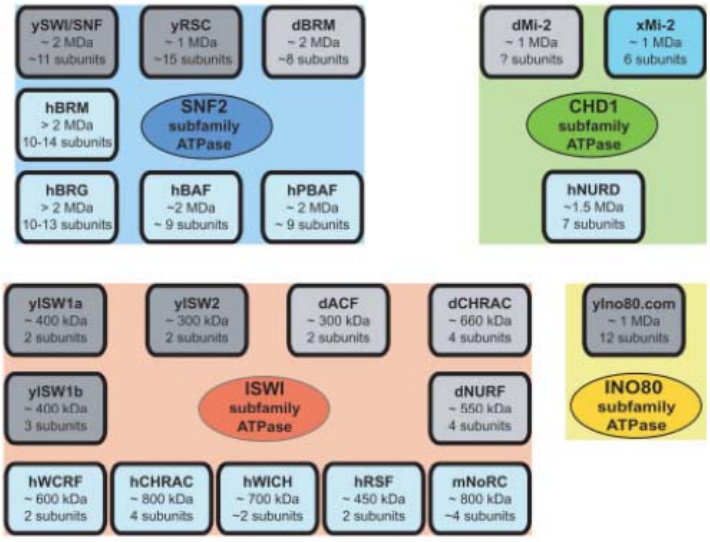

Figure 3- ATP-Dependent Remodeling Factors- This figure lists the SNF2, ISWI, and CHD1 family of ATPases and their subfamilies. The rectangles symbolize chromatin-remodeling complexes. Included are the approximate size of the complex and the number of subunits (Lusser and Kadonaga, 2003).

Swi2/Snf2 Family of ATPases

The Swi2/Snf2 family of chromatin remodelers was discovered through genetic screens in yeast. The family was named for *swi*tch (swi) and *s*ucrose *n*on*f*ermenting (snf) mutations identified within the complexes (Becker and Hörz, 2002). The 11-subunit SWI/SNF complex binds the H2A/Htz1 histone (Krogan, et.al, 2003). This complex increases the accessibility of nucleosomal DNA in an ATP-dependent manner, and is controlled by this remodeling family (Becker and Hörz, 2002). More specifically, Htz1 associates with SWR1 of the Snf2 family ATPase and catalyzes an ATP-dependent histone exchange (Korber and Hörz, 2004). Mutations in the Snf5 subunit (found in higher eukaryotes) showed involvement of the complex in assembly and catalytic functions. Within this family of ATPases, actin-related proteins found to be contained in yeast and higher eukaryotes may be associated with nuclear matrix structures.

While much of the data in relation to this family of ATPases has been observed in yeast, at least two homologs of Snf2 exist in human cells. One of these, brahma (BRM), has also been discovered in *Drosophila melanogaster* and is a gene regulator as part of a multisubunit complex (Becker and Hörz, 2002). In *D. melanogaster*, expression of a dominant-negative protein variant of BRM caused decreased cell viability, defects in the peripheral nervous system and homeotic transformations. The BRM protein has also been found to have high levels of expression in the nuclei of developing organisms (Becker and Hörz, 2002). These findings in *the D. melanogaster* species are hypothesized to parallel functions in humans.

ISWI/SNF2L ATPases

The ISWI/SNF2L family of ATPases were discovered in *D. melanogaster* because of homology to the brahma ATPase domain. The family is named for *i*mitation *swi*tch (ISWI), however it is not related to the Swi2/Snf2 family. In vitro models have shown that ISWI; remodeling factors can stimulate transcription from chromatin templates (Becker and Hörz, 2002). The ISWI ATPase itself has been shown, in vitro, to carry out nucleosome remodeling reactions while associated with other subunits forming numerous distinct complexes. These separate, associated subunits may contribute to ISWI regulation, with increased activity when ISWI is associated into remodeling factors (Becker and Hörz, 2002). The importance of the processes in development is demonstrated by the observation that a homozygous null mutation of ISWI/SNF2L in flies is lethal, with development proceeding only until the late larval stage.

Three ISWI-containing remodeling complexes have been identified in *D. melanogaster* that induce nucleosome sliding, making DNA segments accessible to interacting factors. Two proteins related to ISWI in yeast (Isw1 and Isw2) have been

found to reside in complexes associated with nucleosome remodeling and spacing (Becker and Hörz, 2002). In addition, study of ISWI-containing complexes in *Xenopus laevis* suggest ISWI may be able to disrupt interactions between some nonhistone proteins with DNA.

The CHD Family(NURD Complex)

The NURD complex is part of the CHD family of nucleosome remodeling ATPases. Members of the CHD family are characterized by the presence of a pair of chromodomains, hence the name of the group(Becker and Hörz, 2002). Within the CHD family, are four known specific members, CHD1-4. CHD1 is a DNA-binding protein found in decondensed active chromatin. CHD2 is highly related to CHD1, but little is known about either of their exact roles. More is known about CHD3 and CHD4, also termed Mi-2α or Mi-2β respectively. Mi-2 has been found in the NURD (*nu*cleosome *r*emodeling and *d*eacetylation) complexes of several species. The NURD complexes combine covalent modification, in this case deacetylation, and ATP-dependent remodeling (Becker and Hörz, 2002). One example of the purpose of the NURD complex is its association with the repression of certain cells during development through targeted deacetylation of specific genes in *Caenorhabditis elegans*.

Other Remodeling Factors

Other nucleosome remodeling factors include the INO80 complex and the Cockayne Syndrome B factor. The result of modifications by these complexes is known, but little else is known about the exact function or mechanism of the remodeling factors. In yeast, it has been found that the deletion of the INO80 complex results in an increased sensitivity to DNA damage (Becker and Hörz, 2002). INO80 exists in a complex with Rvb and Rvb2 which are homologs to bacterial RuvB helicase. Also a mutation in the

Cockayne Syndrome B Factor nucleosome remodeling ATPase has been found to cause increased ultra violet sensitivity and abnormal neural developmental in patients (Becker and Hörz, 2002).

Histone Modifications

The roles of assorted histone modifications (Figure 4) are important to study because they affect gene control and most likely, various other biological control elements. Research shows that histone modifications directly affect chromatin structure and allow for surface protein interaction(Berger, 2002). Each modification, whether acetylation, phosphorylation, ubiquination or methylation, correlates with a specific transcription state. Also, individual histone modifications may act alone, or in conjunction with other modifications. The following sections outline the different histone modifications and their hypothesized affects on chromatin structure.

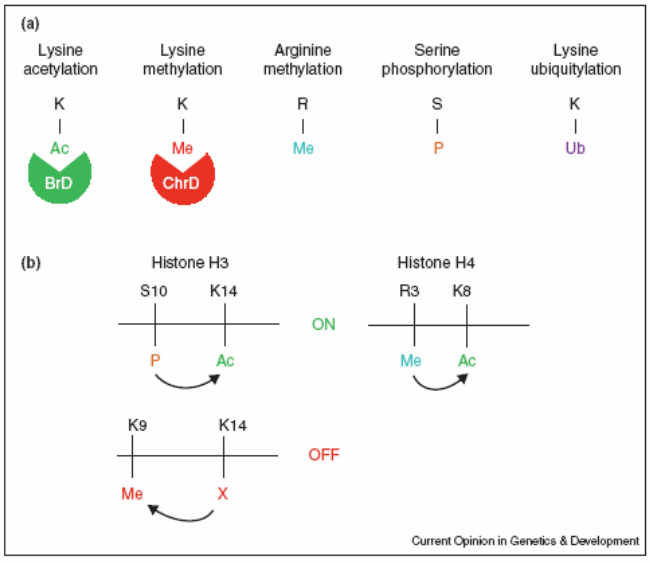

Figure 4- Types of Histone Modifications- (A) Types of modifications include Lys acetylation, Ser phosphorylation, Arg and Lys Methylation, and Lys ubiquitylation. Also shown are the two types of domains that interact in acetylation (bromodomain)and methylation (chromodomain). (B) Patterns/pairs of modifications that correlate with either active or repressed transcription (Berger, 2002).

Histone acetylation, phosphorylation, methylation and ubiquitylation all have been found to correlate with either gene activation or repression. The histones include H1, H2A (H2A.X and H2A.Z), H2B, H3, H4. Most of the modifications are targeted to the N-terminal tails of histones H1-H3, which are structurally more accessible than the others (Kao and Osley, 2003). Modifications are specific to certain amino acid residues, but one modification can associate with a specific residue at numerous locations along the DNA. Individual or sequentially related modifications may dictate specific genomic states, including gene activation, repression, DNA repair, recombination, or chromosome segregation (Berger, 2002). One modification can also have multiple implications on chromatin.

Modifications on histone tails could serve as "tags" and be part of the hypothesized "histone code." This hypothesis includes the theory in which the modification "tags" serve to increase the readout of upstream signaling pathways, causing greater changes in the chromatin structure of target genes (Strahl and Allis, 2000). This reinforces the idea that multiple histone modifications can act together, possibly by influencing the rate or efficiency of subsequent modifications (Strahl and Allis, 2000).

The most well-known and characterized modification is histone acetylation, which has been linked to transcriptional activation and less significantly, histone deposition. Acetylation on lysine 9 or lysine 14, primarily on histone H3, causes the transcriptional activation (Loury and Sassone-Corsi, 2003). Acetylated lysines have also been found to associate with co-activator bromo-domains which recognize the acetylation in the histones (Kouzarides, 2002). Also, because acetylases mediate transcriptional activation, transcriptional repression is therefore carried out by deacetylases which reverse the reaction (Kouzarides, 2002).

Though less is known about it, histone phosphorylation is known to be involved in a wider range of cellular processes than histone acetylation. This is most likely because all core histones contain phosphoacceptor sites on their N-terminal tails where the majority of histone modifications take place (Loury and Sassone-Corsi, 2003). The biological processes affected by histone phosphorylation may include gene activation (serine 10), mitosis and chromosome condensation (serine 10 and serine 28 on H3), apoptosis (serine 139 on H2A.X; serine 32 on H2B; serine 10 on H3, H1 and H4) and damage repair (serine 139 on H2A.X with modification spanning through the bases around the DNA lesion) (Loury and Sassone-Corsi, 2003). Phosphorylation at Ser10 on Histone H3 is the most well documented modification, causing both transcriptional activation/regulation (Berger, 2002) and chromosome condensation and segregation during mitosis (Loury and Sassone-Corsi, 2003).

Histone methylation involves two different types, one targeting arginine residues, and the other targeting lysine residues. Arginine methylation is involved in gene activation in which methylases are recruited to promoters as co-activators. Here, the co-activators target regions on either H3 or H4 (Berger, 2002). Methylation of arginine residues correlates to an active transcription state much the same as acetylation (Kouzarides, 2002). Conversely, methylation of lysine residues varies significantly from arginine methylation in that it is associated with transcriptional silencing, such as at lysine 9 (Loury and Sassone-Corsi, 2003). Lysine 79 in H3 is known to be hypomethylated in a variety of eukaryotic species. Dot1 (a histone methylase that lacks a SET domain) mediated methylation at this site has been found to play an important role in heterochromatic silencing in yeast (Hui Ng, et. al, 2003). Methylation at this site is also proposed to be involved in position effect variegation (Hui Ng, et. al, 2003).

Methylation may also be involved in other non-transcriptional chromatin processes such as DNA repair and recombination (Kouzarides, 2002).

Histone ubiquitination is important in mitotic and meiotic growth, it is still unknown if this modification plays a role in transcription regulation (Berger, 2002). Ubiquitulation is different from other modifications in that ubiquitin is a large molecule, a 76 amino acid protein. It attaches to histones through an isopeptide linkage between its terminal glycine and an ε amino group lysine in the histone (Kao and Osley, 2003). Like acetylation, ubiquitylation is a reversible histone modification. In fact, sites that have been monoubiquitylated *in vivo* display a high turnover rate between ubiquitylation and deubiquitylation (global cycle) during mitosis. Histones that are ubiquitylated are found to be the most stable constituents of nucleosomes in chromatin isolated from vertebrate and fly cells (Kao and Osley, 2003). Like the other histone modifications, ubiquitylation may also be involved in other processes such as remodeling, transcription, repair and replication factors (Kao and Osley, 2003), but these are not yet conclusively proven.

Little else is known of the exact effects of histone ubiquitylation on structure, but it may act as a recognition element, directing proteins to specific chromatin domains. Ubiquitylation, along with other modifications, reinforces the idea of the "histone code" (Kao and Osley, 2003). The "histone code" is proposed to be a pattern of histone modifications along DNA. These modifications are then deciphered by proteins that bind to the chromatin and carry out regulatory decisions (Bulger, et. al, 2002).

Matrix/Scaffold Attachment Regions

Matrix or scaffold attachment regions (termed MARs and SARs respectively) have been of much interest to researchers over the past decade. Rollini et. al (1999)

defines MARs as fragments of genomic DNA with the ability to bind to isolated nuclear matrices in vitro. Found in eukaryotic genomes, MAR elements are hypothesized to play roles in the organization of chromatin structure during interphase and metaphase (Rollini, et. al, 1999).

MARs are often found at the boundaries of transcriptionally active domains, supporting the idea that their amplifying affect on transcription rates are due to insulation against the effects of surrounding chromatin (Phi-Van and Strätling, 1996). Recently, human MARs in engineered *D. melanogaster* have been found to function as insulator elements, shielding integrated transgenes from chromosomal position effects in vivo (Rollini, et. al, 1999). This is due to the secondary structure of MARs, where loops of DNA protect its own DNA from outside influence. MARs have also been found in promoter regions and introns. In these cases, the MARs function in transcription and gene regulation in association with enhancers. Further still, MARs have been found to possibly stimulate expression of heterologous reporters in stably transfected cells (Rollini, et. al, 1999).

The following sequence motifs have been found to be highly enriched in DNA qualifying as MAR and containing binding activity: topoisomerase II binding sites, DNA unwinding motifs, simple sequence motifs (A, T, and H boxes), MAR recognition sequences (MRS), base unpairing regions (BUR), 90% AT repeating regions, and curved and kinked regions. AT rich DNA segments have been found to be the most often occurring motif (Tolstong et. al, 2001). These include long AT tracts with alternating A_n and T_n runs identified by the MAR-binding protein SATB1 (Tolstong et al, 2001). Rollini et. al (1999) used a serine protease inhibitor to study the regulation of gene activity and chromatin structure on human chromosome 14q32.1. They used assays to test where

MARs were inserted, and found that though a common feature of MAR structure consists of repetitive DNA, no specific repeat leads to matrix binding activity.

Matrix attachment region's effects on expression and position effects in CHO cells are the reason so much interest has been placed on them. The Phi-Van and Strätling (1996) literature and their preceding research on the chicken lysozyme gene 5' MAR was the basis for this project. Previous literature had shown that flanking 5' MARs in stably transfected cells enhanced the expression of a reporter transgene and showed decreased position effects of the chromatin structure (Phi-Van and Strätling, 1996). A 6.1 kb enhancer was flanked by lysozyme 5' MARs on each side, and gene expression increased 10 fold due to the boundary sequences (Phi-Van and Strätling, 1996). The MAR location at domain boundaries encouraged the idea that they have an effect on gene expression by an insulating effect. In the previous study by Phi-Van and Stratling, they used a 2.95 kb MAR in cultured cells and found increased expression. In this study they wanted to duplicate their results of increased expression using smaller fragments. Their conclusions were that 5' MAR fragments of 1.32 and 1.45 kb retained the ability to stimulate transgene expression and reduce variation in expression level (Phi-Van and Strätling, 1996). Zahn-Zabal et. al (2001) found that two flanking MARs have a greater effect than one lone MAR when present in the expression plasmid. Two MARs in this study also elevated transgene expression, and continuted to have higher expression (20-fold over the control) over 6 months (Zahn-Zabal, et. al, 2001). Another similar study with insertion of sequences containing MARs yielded 7-fold higher expression over controls (Kim, et. al, 2003).

Shorter fragments that had been tested for having the same affects as the larger MARs were found to be inactive and did not have the same insulating or expression

affects (Phi-Van and Strätling, 1996). It was also found that the ability of MAR fragments to bind to the nuclear matrix is not sufficient to enhance expression and insulation in stably transfected cells. The functionality of the MARs involved other aspects than just an affinity for the nuclear matrix (Kim et. al, 2003). Therefore, the site of the chicken lysozyme is location dependent (Phi-Van and Strätling, 1996).

Removal of the MARs in the cell line using the 1.32 and 1.45 kb fragments decreased transcription between 35-1000 fold. Also, there was an increase in the variation in expression among the different lines within the study (Phi-Van and Strätling, 1996). This finding supports the idea that MARs not only increase transcription levels, but decrease variation in the levels of expression. The variation in expression levels, which is verifiably decreased in sequences containing MARs, is due to the position effect of chromatin at the sight of integration into the host genome (Zahn-Zabal, et. al, 2001).

MAR elements from other genes and species, primarily plants, have been found to have insulating and transcription enhancing activities comparible to those of the chicken lysozyme 5' MAR in the experiments from Phi-Van and Stratling (1996). This is encouraging information in the further study of MAR regions, and supports the study of this MQP project.

Previous Work at Abbott Related to this Project

Study into MAR "genomic sequences" began in May, 2000. Much work had already been done prior to the beginning of this portion of the project. The originally identified high expressing clone's genomic phage library was constructed and screened. From here the flanking "genomic" regions were isolated and sequenced. The "genomic arms" were named p3.8AM13R (2.422 kb) and 2101EX (2.361 kb). Figure 5 shows the distribution of MAR elements within both of these genomic arms.

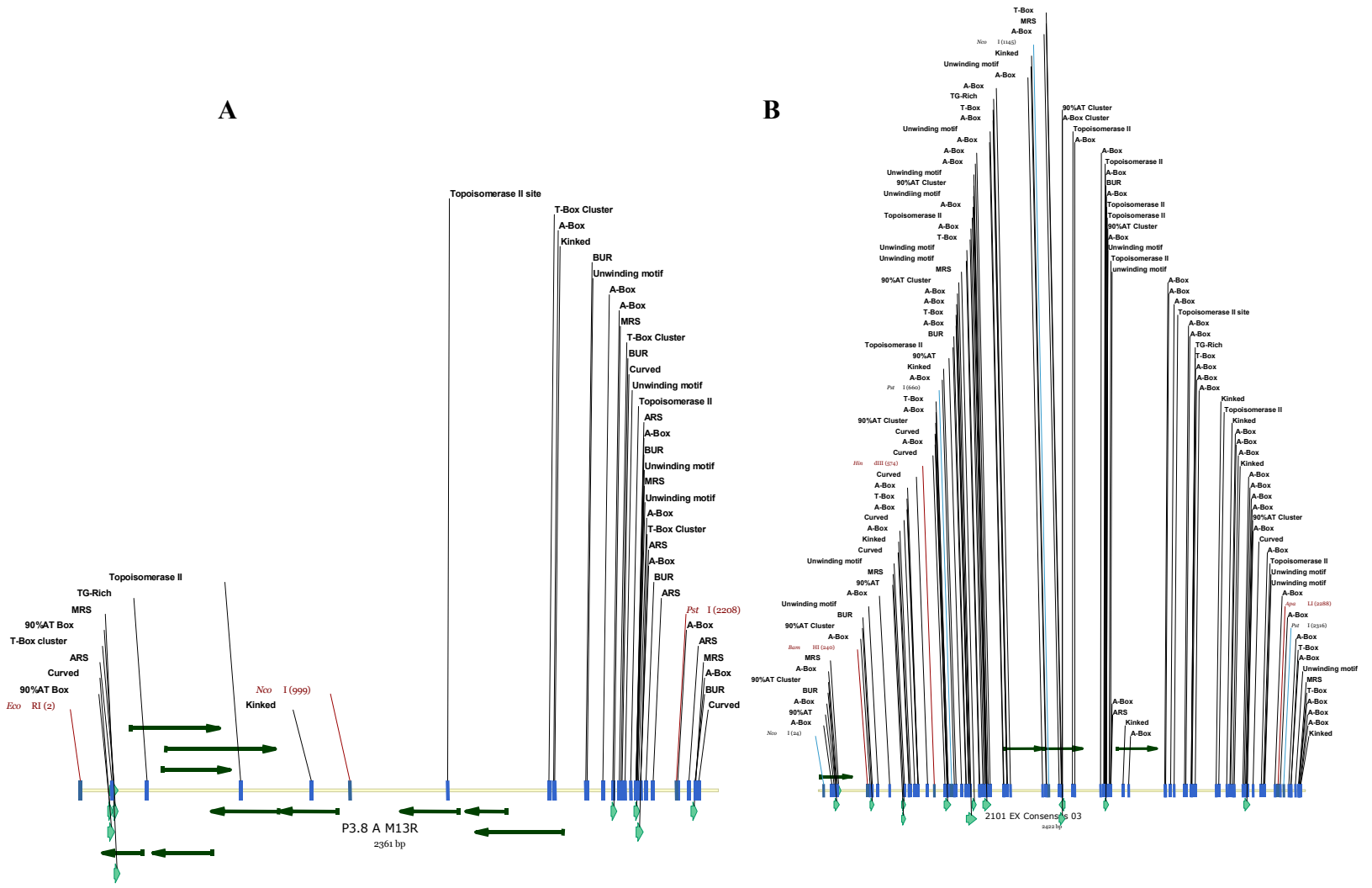

Figure 5- Distribution of MAR Elements Within the Abbott Cloned Genomic Arms.- (A) Location of MAR elements in the P3.8AM13R genomic arm. **(B)** Location of MAR elements in the 2101EX genomic arm. These portions are shown in blue on the plasmid and named with the appropriate motif.

To construct the genomic vector, the genomic arms were inserted into another one of Abbott's previously studied plasmids. The decision to use this particular vector included previous difficulties in getting consistently high expression in its cell lines, and the need to test the genomic arms with a different antibody to solve the question of whether the increased expression was specific to only the original antibody producing clone.

Each of the genomic arms was subcloned from separate plasmids into one resultant plasmid, pA205. The P3.8AM13R segment was cloned from pA204 to create

pA205 E/E insert. The 2101EX genomic segment was cloned from pA190 to create pA190 P/P insert. The two inserts were then ligated together to achieve the final pA205genomic plasmid.

PROJECT PURPOSE

Researchers at Abbott Bioresearch Center have identified an antibody producing cell line that expresses higher levels of antibody than those previously studied. In this line, antibody levels increased between 40%-150% in the bioreactor compared to the average produced by this cell line. Specific genomic segments were identified within the clone to be responsible for increased protein expression and hypothesized to be due to the genomic integration site.

Abbott's goals were to find if the sequences could confer higher expression into other expression vectors, to identify which specific portions within these genomic regions were responsible for the increased expression, and to determine if these were in fact matrix attachment regions. These questions could be approached by setting up a system to study the structure of the possible MAR sequences. To do this, the two possible MAR sequences were cloned into a different antibody vector along with a Green Fluorescent Protein (EGFP) as a reporter gene, in place of the antibody coding region. The variable plasmid and the control parent plasmid were transfected into Chinese hamster ovary (CHO) cells and tested using FACS analysis. The fluorescence was used to determine positive clones, and gene expression levels. If these regions are found to increase expression, this information will be used to design next generation vectors, leading to more robust CHO cell line development processes at Abbott Laboratories.

METHODS

Cloning of Parent and Genomic Plasmids

The cloning of the parent and genomic plasmid containing the possible MAR regions included two attempts before the plasmids were finally able to be constructed. Both final constructs included removal of the light chain and heavy chain antibody sequences and insertion of the EGFP gene, which would later be used as a fluorescence detection mechanism. The major obstacle encountered while trying to achieve successful cloning was the size of the genomic plasmid. The original size of pA205genomic was 14140 bp, and the final clone size is 10201 bp. The attempts to obtain the experimental genomic and parent plasmids included a PCR amplification method and a multi-step process to eliminate the antibody sequences.

Strategy 1

The first strategy of restriction enzyme digests of the parent pA205 and pA205genomic plasmids consisted of sequential digests of with *Bsi* WI and *Kpn* I. These enzymes were chosen because of the site location, eliminating the light chain and heavy chain sequences of the antibody from the vectors, and only one digestion site for each, eliminating partial digestions, and also to achieve complementary ends for ligating to the EGFP insert. After digestion, with *Bsi* WI and *Kpn* I, both samples were run on a 0.8% agarose gel to separate the fragments. The correct bands were excised from the gel and purified using the QIAquick Gel Extraction Kit following protocols recommended by the manufacturer to obtain the purified DNA fragments for each sample. The pEGFP plasmid was digested with Kpn I and *Apa* I simultaneously, and the 831bp fragment insert was isolated as outlined above.

Next, the fragments from parent and genomic underwent polymerase chain reaction (PCR) to amplify the products. Three different types of PCR were attempted and also cycling changes were made to try to obtain the highest yield of DNA product. These included Elongase Enzyme Mix and Protocol (as suggested by the Invitrogen), Supermix PCR, and Pfu Turbo PCR mix. The Elongase method required determining differing amounts of 2 buffers to obtain the best yield of product. For each PCR run, at least 10 tubes were set up for either the parent or genomic plasmids. After PCR, both of the products from the parent and genomic were ligated to the EGFP gene insert to obtain the intended final plasmid constructs.

After the second PCR run, Supermix PCR was identified as giving the highest yield of product, so Supermix was used in all subsequent PCR runs. For PCR using Supermix, 45 µl Supermix, 2 µl of each primer, and 1 µl of the purified template (either parent or genomic) are added into one PCR tube and run overnight using the Elongase cycling. After each PCR run, 5 µl of samples were run on an agarose gel to determine if PCR had yielded the correct products, or any product at all. Tubes containing correct products were pooled and run on a gel to separate out any incorrect fragments. The correct bands were excised from the gel, gel purified, and ligated to the EGFP insert fragment.

The first three runs of PCR yielded the correct products for both the genomic and parent fragment, and each time the samples were pooled and ligated to EGFP using the Rapid Ligation buffer. However, few colonies grew on the LB + Ampicillin plates, and all were determined to be incorrect transformants (as determined after miniprep plasmid purification using the Promega Wizard Prep Kit). The next three runs of PCR produced no products, and the final attempt yielded only incorrect products. It could not be

determined why the correct fragments were unable to be amplified using PCR. Because of the failed attempts to achieving either the parent or genomic plasmids using this method, an alternative strategy was devised.

Strategy 2

The second strategy to obtain the parent and genomic plasmids with the inserted EGFP gene and removal of the heavy and light chain antibody sequences, involved a multi step process for both. The steps included adding an Nhe I linker to make complementary ends for the EGFP insert ligation, and a separate removal of the heavy and light chain antibody sequences.

For the parent pA205 plasmid (not containing the genomic sequences), construction began by digesting with *Nru* I and *Pst* I simultaneously to obtain the 4791 bp band. This digestion removed the heavy chain antibody sequences. Next, parent pA205 was ligated to a *Nhe* I linker from New England Biolabs. The *Nhe* I linker will make complementary ends on the parent for later ligation to the EGFP gene (digested with *Xba* I). The parent was then linearized by digestion with *Nhe* I. pEGFP was cut with *Xba* I to isolate the EGFP gene, the 763 bp fragment. Next, the parent pA205-Nhe I linker was ligated to EGFP to obtain the intermediate construct, parent pA205-light chain-EGFP. This intermediate was then digested with *Not* I to remove the light chain sequence (669 bp fragment). Finally, the fragment was self ligated to obtain the final constructed clone: parent pA205-EGFP.

The pA205 genomic plasmid was digested with *Nru* I to obtain the 9430 bp band (heavy chain removed). This band was isolated and ligated to a *Nhe* I linker from New England Biolabs. After ligation, pA205genomic was cut with *Nhe* I to linearize. After several different digestions with incorrect results, Calf Intestinal Alkaline Phosphatase

(CIP) from New England Biolabs was also used. After digestion with *Nhe* I to linearize, CIP was added to to remove 5' phosphates from the plasmid to prevent self-ligation and recircularization before EGFP is ligated into the vector. The pEGFP plasmid was cut with *Xba* I, and the 763 bp band containing the EGFP gene was isolated. Next, pA205genomic-Nhe I linker was ligated to EGFP to obtain the pA205-EGFP-light chain intermediate. Next, the intermediate plasmid was partially digested with *Not* I, and the 9532 bp band was isolated and purified. This digest eliminates the light chain antibody sequence. Finally the linear fragment was self ligated to obtain the final constructed clone: pA205genomic-EGFP.

Cell Culture: Transient Transfection: Fluorescence

The first transfection performed was a transient transfection to test fluorescence. The transfection was done with B3.2 CHO cells using the $CaPO_4$ transfection protocol with pA205genomic-EGFP clone DNA. The purpose of this transfection was to check whether the cells had taken up the plasmid and the EGFP is producing fluorescence as predicted. The transfection consisted of 3 10cm tissue culture plates, which were incubated at 37^{o}C for 2 days. After 2 days, the plates were viewed and photographed using a fluorescence microscope.

Cell Culture: Stable Transfection: FACS

Next, stable transfections were done, also using the $CaPO_4$ protocol (protocol from Jackie Welles). These transfections were done to test if the pA205genomic-EGFP was producing higher transfection rates and more stable clones. If this was indeed the case, pA205genomic-EGFP transfected cells would be producing more fluorescence versus the parent pA205-EGFP control cells. Also included in the study was the parent

pA205 (-) as a negative control, and the hCMV-EGFP positive control. The parent pA205 (-) contained no EGFP so should show no fluorescence, and the hCMV-EGFP is a high expressing positive control with very high fluorescence. Selection of the cells was done using hygromycin. A hygromycin resistance gene was transfected into the cells at the same time as the respective DNA. After day 2, hygromycin was added to the feed media.

Three stable transfections were performed and sorted, with multiple sorting cycles. The first began October 12th, 2004 and was sorted on October 28th (Sort 1). All of the cells were in T-150 flasks and had been fed with α-MEM + 5% FBS 1xHT and 400 μg/mL hygromycin. The next transfection began November 4th and was sorted on November 30th (Sort 2). These cells had been fed with media containing 250 μg/mL hygromycin. The sorted cells from Sort 2 were resorted on December 8th, 2004 and during this period had been changed to media containing 400 μg/mL hygromycin. The final stable transfection began on December 4th, 2004 and was sorted on December 21st (Sort 3). During this transfection, the cells had been fed only with media containing 400 μg/mL hygromycin. Sort 3 was resorted on January 27th, 2005. All of the FACS analysis was done on the Moflo FACS machine by Sukumari Mohan.

RESULTS

The first step in investigating the elevated expression levels of the original clone in the production cell line, was to construct a plasmid clone containing both genomic arms present in the clone flanking an EGFP reporter gene. Plasmid pA205-EGFP (Figure 7A) containing EGFP served as the parent control plasmid. Two genomic regions, P3.8AM13R and 2101EX were subcloned into this parent plasmid to create plasmid pA205genomic-EGFP (Figure 7B).

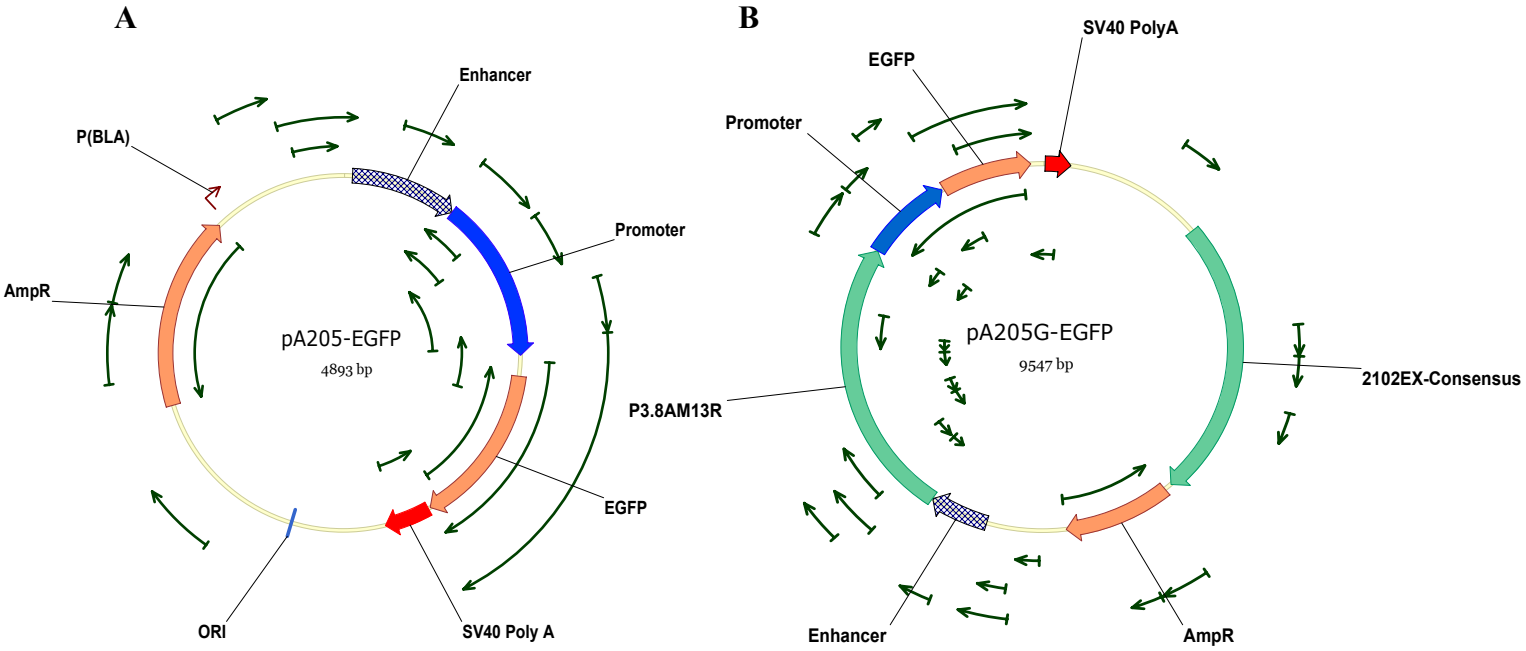

Figure 7- Final Plasmid Constructs- Panel **(A)** denotes parent vector pA205-EGFP. Panel **(B)** is the final clone of pA205genomic-EGFP containing both genomic arms labeled: P3.8AM13R and 2101EX.

Next, the transient transfection experiment was done to verify that the inserted EGFP gene was capable of being expressed, and the cells had taken up the pA205genomic-EGFP DNA. Figure 8 shows the fluorescence microscopy pictures taken of the three transfected plates.

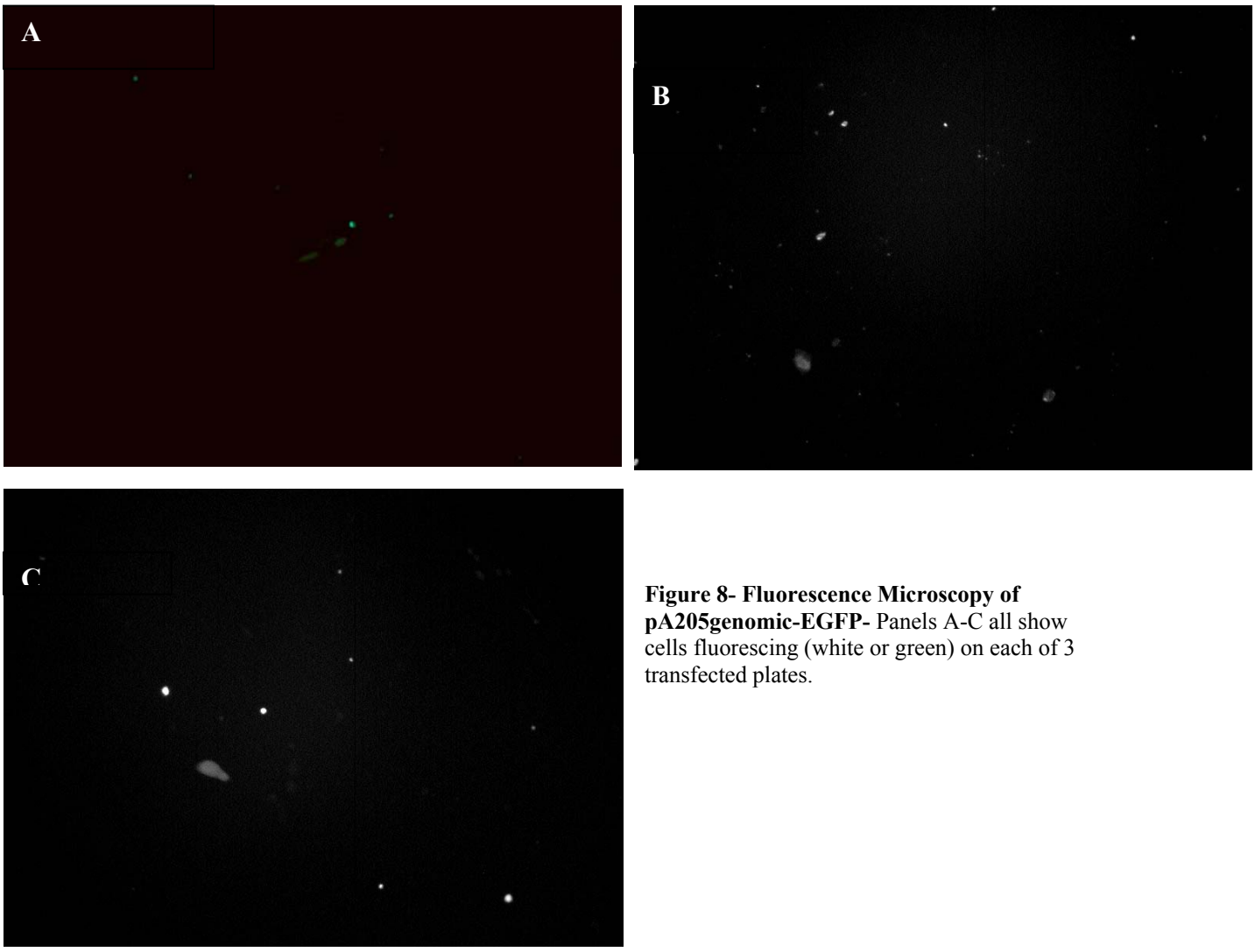

Figure 8- Fluorescence Microscopy of pA205genomic-EGFP- Panels A-C all show cells fluorescing (white or green) on each of 3 transfected plates.

Although fluorescence was detected, not all of the cells are fluorescing. The plates were nearly confluent when these pictures were taken. This observation is expected in that not all cells will take up the plasmid DNA.

Once it was verified that EGFP was capable of being transfected and expressed, stable transfections were performed. For each transfection, FACS analysis was done on the cells for parent pA205-EGFP, pA205genomic-EGFP, parent pA205(-) control, and hCMV-EGFP(+) control to determine whether the MAR sequences in the genomic arms of pA205genomic-EGFP were increasing transfection and stability in these cells by

showing a shift in fluorescence. Table 1 outlines some of the results from all of the FACS sorts.

Date of Sort	10/28/05	11/30/04	12/8/04	12/21/04	1/27/05
Resort			of 11/30 sort		of 12/21 sort
Flask	T-150	T-75	T-75	T-75	T-150
Concentration of Hygromycin (ug/mL)	400	250	400	400	400
Ratio of sorted cells (genomic:parent)	5455:2225	83097:79836	358719: 102341	45424:97259	51060:52797
Results	slight shift in fluorescence from parent	slight shift in fluorescence from parent	no shift in fluorescence	slight shift in fluorescence from parent	distinct shift in an increase in fluorescence from parent

Table 1- FACS Analysis Results of p205genomic-GFP- Each of the sorts are outlined including information on whether it was a resort of a previous sort, the size of the flask, concentration of Hygromycin, the number of parent and genomic cells that were sorted, and a note of the overall result.

The following figures are histograms of fluorescence versus cell count for each of the FACS analyses. Parameters of the sorts were based on the parent pA205 (-) control to set a basis for background fluorescence. With the negative fluorescence set, any signal above this point represents positive fluorescence and is detected by the Moflo equipment. In the figures, the (-) control histograms are not shown, this is because it was only used as a baseline and no histograms were made. None of the parent (-) controls ever expressed fluorescence due to the lack of EGFP gene in the plasmid. Figure 9 contains the results from the first sort done on 10-28-04.

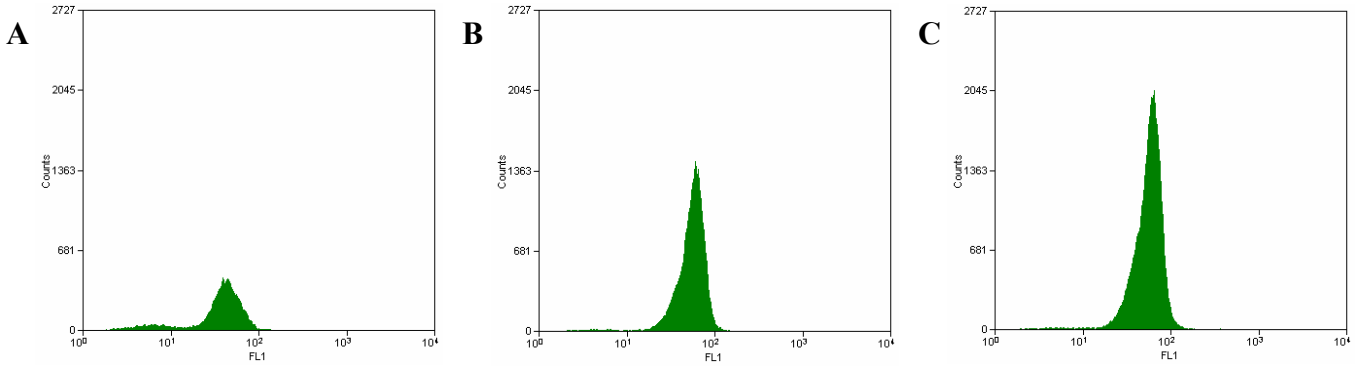

Figure 9- October 28, 2004 FACS Data- (A) Fluorescence of hCMV-GFP (+) control, **(B)** represents parent pA205-GFP, and **(C)** denotes pA205genomic-GFP. The hCMV-GFP + control shows little to know level of fluorescence due to the low cell counts exhibited by all of the samples.

In Figure 9, a small hump is visible in pA205genomic-GFP (panel C) that shows a slightly higher level of fluorescence than in parent pA205-GFP (panel B). However, the slight increase is not great enough to show a definite increase in cells fluorescing over the parent, caused by MAR regions.

To repeat the results, or get even better results than those from the first sort, a second transfection was performed, and sorted on November 30, 2004. Figure 10 shows histograms of the FACS results from the second sort on November 30, 2004.

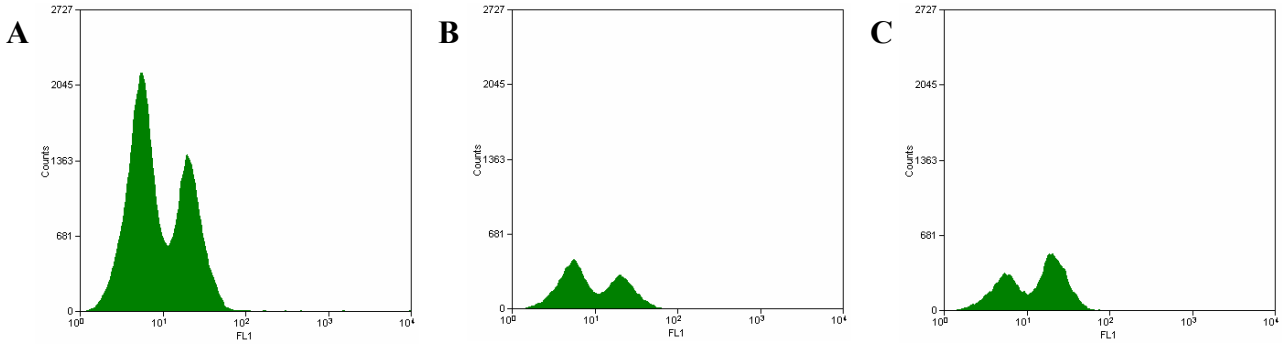

Figure 10- November 30, 2004 FACS Data- (A) Fluorescence of hCMV-GFP (+) control, **(B)** represents parent pA205-GFP, and **(C)** denotes pA205genomic-GFP.

In Figure 10, there is no change in fluorescence between the cells containing the parent control plasmid (panel B), and the cells containing the MAR segments (panel C). However, it appears that the percentage of cells that display high fluorescence is greater

for the pA205genomic vector. A resort was done on these sorted cells on December 8, 2004. Figure 11 shows the results from this resort on the same cells.

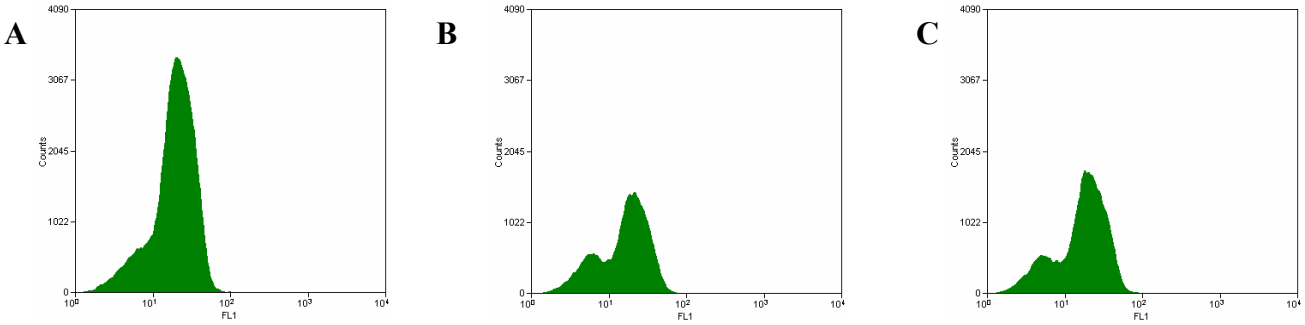

Figure 11- FACS Resort of November 30, 2004 Data – (A) Fluorescence of hCMV-GFP (+) control, **(B)** represents parent pA205-GFP, and **(C)** denotes pA205genomic-GFP.

The resort on the November 30, 2004 sort showed little or no difference in maximum fluorescence between the parent pA205 and pA205genomic cells. However, the percentage of transfected cells displaying maximum fluorescence was again greater for the latter vector (Figure 11). A third transfection was performed, and FACS analysis was done on December 21, 2004. Figure 12 shows the results from this transfection's FACS sort.

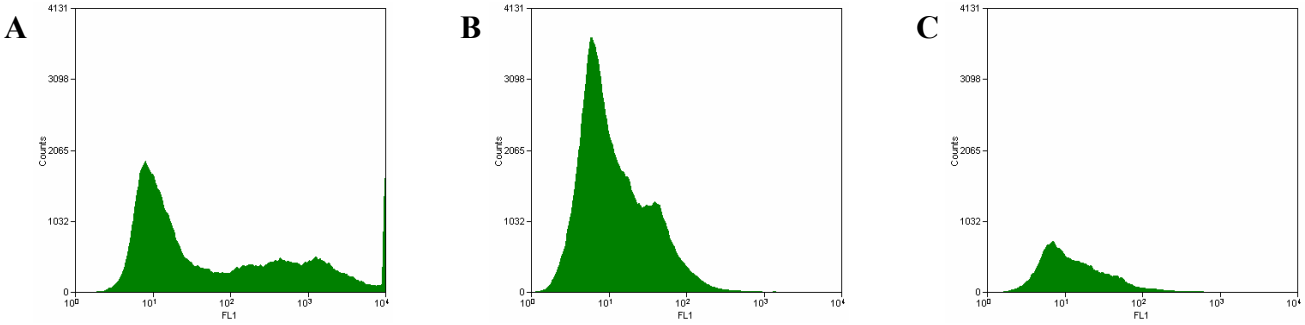

Figure 12- December 21, 2004 FACS Data - (A) Fluorescence of hCMV-GFP (+) control, **(B)** denotes parent pA205-GFP, and **(C)** represents pA205genomic-GFP.

The data from the third transfection shows a slight shift to the right in the cells containing the MAR sequences, indicating increased fluorescence. However, the number of cells

varies greatly between the parent and genomic samples, so it is the data is inconclusive on how much the fluorescence increases. To obtain more definite results, a resort of this transfection was done on January 27, 2005 (Figure 13).

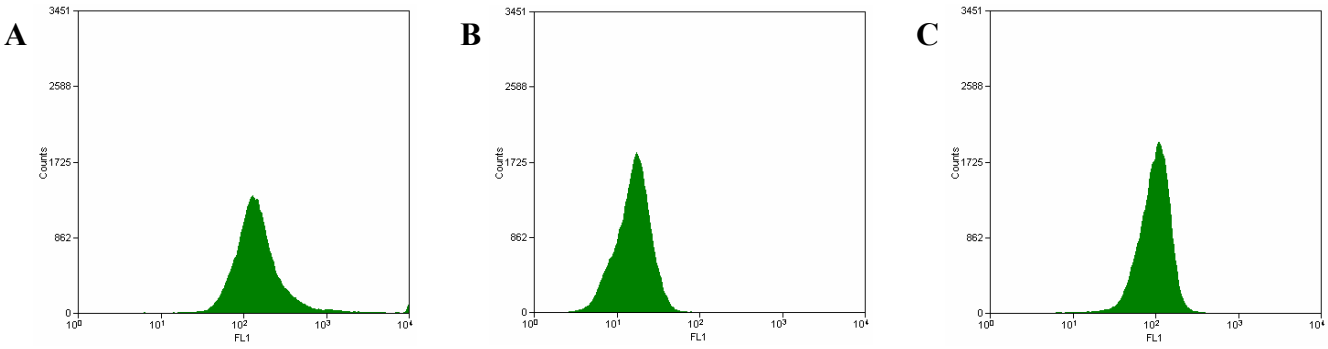

Figure 13 – FACS Resort of December 21, 2004 Data – (A) Fluorescence of hCMV-GFP (+) control, **(B)** denotes parent pA205-GFP, and **(C)** represents pA205genomic-GFP.

The resort on the third transfection shows a definite shift to the right in the level of fluorescence (panel C relative to panel B). The number of cells sorted in this FACS run was nearly equal, whereas in all of the previous data, the number of cells was not adjusted to be equal. This may have played a role in the significant increase in the level of fluorescence observed from this data.

DISCUSSION

Before analyzing the results of this study, it is beneficial to briefly review them. First, experimental plasmids were constructed by removing the light chain and heavy chain antibody sequences of a known high expression clone. In the "genomic" plasmid, supposed matrix attachment regions had been ligated in to test their MAR activity. Flanking MAR fragments of 1.32 kb and 1.45 kb have been previously tested and are known to produce MAR activity (Phi-Van and Strätling, 1996). The MAR fragments tested in this study are both over 2.3 kb, so they are more than long enough to produce activity and have insulating effects. To facilitate analysis of transcription and gene expression controlled by the MARs, and Enhanced Green Flourescence Protein (EGFP) reporter gene was also ligated into the clones. Next, a transient transfection was done with the pA205genomic-EGFP plasmid to verify the activity of EGFP. Finally, a series of stable transfections were done for FACS analysis. These analyses would conclude whether or not the genomic sequences indeed enhanced transcription and gene expression rates, and thus have MAR activity.

The histograms of the samples sorted using FACS were the primary means in deciding whether there was a shift in fluorescence caused by the parent and genomic plasmids. MAR activity would be indicated by an increase in cells fluorescing in the genomic sample (shift to the right) versus the parent. Slight shifts to the right (increasing fluorescence) of pA205genomic-GFP in the first two sets of FACS data produced positive results, but these results were not enough to prompt a conclusive result of MAR activity (Figures 9 and 10). However, the final resort done on January 27, 2005 (Figure 13) produced strong data to conclude that the genomic regions did in fact have a positive

effect on transcription and gene expression amplification. Because of the varying results, it is possible that the addition of the genomic DNA sequences did not always produce cells with significantly higher expression of GFP, but did increase the frequency of successful integration of the functional expression vector (which would also score as an increase in fluorescence). Successful integration would be facilitated by making the chromosomal integration site more conducive to gene expression. This may be the reason for the higher percentage of good expressers from the pA205genomic transfections versus the parent.

There are three possible reasons for the resort of the third transfection producing the best results indicative of MAR activity. First, the third transfection was grown for a longer amount of time (nearly nine weeks) versus all previous transfections (about 2-3 weeks). In stably transfected cells with MAR sequences, research shows that these cell lines are more stable over time (Rollini, et. al, 1999). Second, the final FACS data was based on approximately an equal number of cells, so this gave the most reliable data. In previous sorts, the number of cells had been disregarded and this may have affected histogram appearance. If the number of cells in the parent sample greatly exceeded the number in the genomic sequence, increased fluorescence in the latter cells may not have been as visible as if the sort been based on an equal number of cells. The insertion of MARs increase the proportion of high-producing clones within a cell line (Zahn-Zabal, et. al, 2001). Finally, variation in expression levels is decreased when MAR activity is present (Zahn-Zabal, et. al, 2001). Results from the plasmid containing MAR sequences stayed steady throughout all of the FACS sorts while results from the plasmid without the MAR sequences fluctuated.

Not only should cells containing pA205genomic-GFP be more stable, they should also maintain increased expression levels over periods of time up to six months (Zahn-Zabal, et.al, 2001). Another aspect of MAR activity is position effect. MARs insulate against surrounding chromatin and modifications, but must be long enough to exhibit these affects (Phi-Van and Strätling, 1996). The MARs being tested in this study are both over 2.3 kb, longer than MARs previously shown to have these affects.

The concluding result from this study is that the MAR sequences cloned into pA205genomic produce MAR enhancing activity and are therefore functional. These qualities include increased transcription levels, enhanced gene expression, and cell line stability.

Future experiments on the identified MAR sequences would include conducting repeat transfections, in addition to identifying any specific integral regions. First, another transfection should be performed, as previously done, to reproduce the results obtained in transfection 3. This transfection should be carried out for a longer period of time, up to several months. The data in this study, along with literature, shows that the longer the cell line is grown, the better the MAR activity should be. So if this is true, the next transfection should produce even better results in fluorescence levels.

Further study should include testing of the pA205genomic cell lines to determine if the regions around the integration site form nucleosomes. Also, it should be determined whether the associated histones are susceptible to modifications. However, these studies are outside the scope of this MQP.

Also, defined integral regions within the MAR fragments should be identified. Both of the genomic fragments are very long and hard to work with when cloning into a plasmid vector. If these regions were shorter, or perhaps if there was only one fragment,

it would facilitate cloning and transfection. To do this, one would test MAR enhancing activity after either one of the MAR fragments had been eliminated and also after elimination of small segments of the fragments and truncation of the arms. In the elimination of small segments, transposons could be used by randomly inserting into the plasmid. Then, the plasmid could be digested with the particular restriction enzyme to eliminate whatever segment had been removed. Finally, the plasmid could be self ligated and tested for MAR activity. The only possible drawback to using shorter MAR fragments would be that the average expression level of cell lines increases with the number of MARs present in the plasmid (Zahn-Zabal, et. al, 2001). So the if integral MARs are deleted from the sequence, the prevalence of high producing clones would likely decrease.

BIBLIOGRAPHY

Addison Wesley Longman Publishing (1999) Pearson Education

Access Excellence @ The National Museum of Health (2005) Chromosome. http://www.accessexcellence.org/RC/VL/GG/chromosome.html

Becker PB, Hörz W (2002) ATP-Dependent nucleosome remodeling. *Annu. Rev. Biochem.* 71: 247-273.

Berger S (2002) Histone modifications in transcriptional regulation. *Current Opinion in Genetics & Development.* 12: 142-148.

Boeger H, Griesenbeck, Strattan JS, Kornberg RD (2003) Nucleosomes unfold completely at a transcriptionally active promoter. *Molecular Cell.* 11:1587-1598.

Bulger M, Sawado T, Schübeler D, Groudine M (2002) ChIPs of the β-globin locus: unraveling gene regulation within an active domain. *Current Opinion in Genetics & Development.* 12: 170-177.

Hui Ng H, Ciccone DN, Morshead KB, Oettinger MA, Struhl K (2003) Lysine-79 of histone H3 is hypomethylated at silenced loci in yeast and mammalian cells: A potential mechanism for position-effect variegation. *PNAS.* 100(4): 1820-1825.

Kao C-F, Osley MA (2003) In vivo assays to study histone ubiquitylation. *Methods.* 31: 59-66.

Kim J-M, Kim J-S, Park D-H, Kang HS, Yoon J, Baek K, Yoon Y (2003) Improved recombinant gene expression in CHO cells using matrix attachment regions. *Journal of Biotechnology.* 107: 95-105.

Kouzarides, T (2002) Histone methylation in transcriptional control. *Current Opinion in Genetics & Development.* 12: 198-209.

Korber P, Hörz W (2004) SWRred not shaken: mixing the histones. *Cell.* 117:5-7

Krogan NJ, Keogh M-C, Datta N, Sawa C, Ryan OW, Ding H, Haw RA, Pootoolal J, Tong A, Canadien V, Richards DP, Wu X, Emili A, Hughes TR, Buratowski S, Greenblatt JF (2003) A Snf2 family ATPase complex required for recruitment of the histone H2A variant Htz1. *Molecular Cell.*12: 1565-1576.

Loury R, Sassone-Corsi P (2003) Histone phosphorylation: how to proceed. *Methods.* 31: 40-48.

Lusser A, Kadonaga JT (2003) Chromatin remodeling by ATP-dependent molecular machines. *BioEssay.* 25:1192-1200.

Phi-Van L, Strätling WH (1996) Dissection of the ability of the chicken lysozyme gene 5' matrix attachment region to stimulate transgene expression and to dampen position effects. *Biochemistry*. 35:10735-10742.

Rollini P, Namciu SJ, Marsden MD, Fournier REK (1999) Identification and characterization of nuclear matrix attachment regions in the human serpin gene cluster at 14q32.1. *Nucleic Acids Research.* 27(19): 3779-3791.

Strahl BD, Allis CD (2000) The language of covalent histone modifications. *Nature.* 403: 41-45.

Tolstong GV, Mothes E, Shoeman RL, Traub P (2001) Isolation of SDS-stable complexes of the intermediate filament protein vimentin with repetitive, mobile, nuclear matrix attachment region, and mitochondrial DNA sequence elements from cultured mouse and human fibroblasts. *DNA and Cell Biology.* 20(9): 531-554.

Zahn-Zabal M, Kobr M, Girod P-A, Imhof M, Chatellard P, de Jesus M, Wurm F, Mermod N (2001) Development of stable cell lines for production or regulated expression using matrix attachment regions. *Journal of Biotechnology*. 87: 29-42.