THE CHARACTERIZATION OF MCEF

An AFF Transcription Factor associated with Acute Lymphoblastic Leukemia and HIV-1

by

Zainab Al Shawaf

B.Sc. Honour Biochemistry, McMaster University 2006

A Research Project presented to Ryerson University
in partial fulfillment of the requirement for the degree of

Master of Applied Science
In the program of Environmental Applied Science and Management

Toronto, Ontario, Canada, 2010

Author's Declaration

I hereby declare that I am the sole author of this Project paper.
I authorize Ryerson University to lend this Project paper to other institutions or individuals for the purpose of scholarly research

I further authorize Ryerson University to reproduce this Project paper by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

THE CHARACTERIZATION OF MCEF

An AFF Transcription Factor Associated with Acute Lymphoblastic Leukemia and HIV-1

Master of Applied Science (M.A.Sc.), 2010

Zainab Al Shawaf

Environmental Applied Science and Management

Ryerson University

Abstract

Acute Lymphoblastic Leukemia (ALL) results from environmentally-triggered *in utero* translocations between the Mixed Lineage Leukemia (MLL) gene and partner genes. In the most frequent cases of ALL, this partner gene is one of the AF4 family (AFF) of transcription factors. The newest AFF member to be discovered and cloned is AFF-4/AF5q31/MCEF. MCEF interacts with a transcription factor necessary for transcription of HIV-1. In addition, evidence suggests that male knockout mice are azoospermic. Therefore, the characterization of MCEF is clinically and theoretically important. The purpose of my research was to further characterize MCEF. In this paper, I first review the AFF members, focusing on MCEF. I then show a series of experimental results addressing MCEF isoforms and HIV-1 repression domains, as well as the generation of anti-MCEF antisera. Finally, I highlight intriguing results with live virus replication assays that suggest how MCEF could be exploited as a therapeutic target for AIDS.

Acknowledgment

I want to first make it very clear that all the experiments and intellectual ideas for this paper, were done by several persons and not just by me, although I group them together in this paper. I would like to sincerely thank my faculty supervisor, Dr. M. Estable, for his much appreciated guidance during the three years I worked in his laboratory, as well as his hand in editing and correcting this paper, helping me bring it to its final format. I also acknowledge his hand in the live virus work and in the MCEF repression domain deletion work and his part in all aspects of the intellectual design of the work behind this paper. I also acknowledge Robert Hopewell for his hand in the MCEF repression deletion work with Dr. Estable and Juzer Kakal (University of Ottawa) for providing me with the experimental details he used in the live virus work with Dr. Estable. Very special thanks go to my colleague, who became my cherished friend, Beverlee Buzon, for lending me her unconditional insight, support and assurance (and very occasional laughs) since she joined the lab. I also acknowledge her hand in the phosphatase experiments. Last but not least, I thank my parents and my sister Rabab for their faith and encouragement; my friend Rasha in Los Angeles for spending endless hours with me over the phone to reassure me that things will work out; my friends Fouad and Siba for waiting with me in the hallways of Kerr Hall North during after-hours until I finished my experiments; and my wonderful, patient husband Laith for being my source of strength through this journey.

Table of Contents

List of Tables

List of Figures

Keywords
AFF4/ Af5q31/MCEF; AF4; P-TEFb; HIV-1; MLL; ALL.

1. INTRODUCTION

The term 'environment', when used in the context of human gestation, can apply to both the mothers' immediate surroundings, as well as the *in utero* environment she creates, as a consequence of maternal nutrition, genetics and health. Just as any other environment can affect the future well-being of the life forms it surrounds, the *in utero* environment can have serious consequences on not only fetal-health, but also neonatal-, infant-, child- and even adult-health [1]. Physicians will normally recommend that mothers not subject the fetus to any form of physical or physiological stress to preserve a healthy environment in the uterus. Extensive research has been conducted in the field of *in utero* diseases, their origins and their impact on a child's wellbeing following birth; studies have proven repeatedly that physiological stress can be brought about by unsuitable environmental conditions. Triggers such as tobacco and drugs, among other chemicals, as well as dietary modifications during the gestation period, are thought to be involved in the onset of *in utero* diseases [2-4]. In addition, ionizing radiation has also been identified as a major risk factor in the development of certain cancers *in utero* [5]. Sufficient evidence has been collected so far to establish the importance of certain nutrients for the health of the fetus during pregnancy, including folate, vitamin B_{12} and vitamin D [1, 6, 7]. A majority of the research is focused on the effect of fetal nutrient deficiency on fetal growth rate and body size at birth [1]. Results have suggested that the nutritional environment can influence birth size, which in turn can predispose the infant for the later onset of many malignancies including cardiovascular disease, diabetes and osteoporosis [8-11], as well as abnormalities in many parts (heart, urinary tract, limbs) of the body [2]. This takes place due to a phenomenon known as fetal programming, where the expression of a fetus' genome is thought to be negatively modified by the lack of essential nutrients during periods of cell division and growth [12]. In addition, certain

cases of physiological imbalances result in childhood Acute Lymphoblastic Leukemia (ALL).

1.1 Acute Lymphoblastic Leukemia

In addition to blood plasma, the blood contains blood cells, and they fall into three main types: red blood cells (erythrocytes), white blood cells (leukocytes) and platelets, all of which are generated in the bone marrow. The primary role of erythrocytes is the transport of oxygen to cells, whereas leukocytes constitute the major part of the immune system, leaving platelets to repair damaged blood vessels. Blood cancer occurs when one or several of these blood cells undergo continuous cycles of cell division and uncontrollable proliferation. Many blood cell cancers have been indentified and these include: myeloma, cancer of plasma cells, which are white blood cells that secrete antibodies; lymphoma, cancer of lymphocyte white blood cells that activate T cells and also generate antibodies; and leukemia. It has been established that all three types of blood cell cancers can arise *in utero* [13-16]. ALL is one of the many forms of blood cell cancers. This type of leukemia is characterized by an excess of lymphoblasts, which normally differentiate into lymphocytes, the white blood cells that protect the body against infections and diseases [17]. In ALL, lymphoblasts produced in the bone marrow, develop mutations that render them capable of multiplying uncontrollably, all the while unable to differentiate into normal lymphocytes, leaving the body weak and highly susceptible to infections [17]. ALL has been identified as the most common cancer (representing 30%), and most common leukemia (80%) in infants and children [18, 19] under the age of 15 (frequency peaks between 2 and 5 years of age). It has been estimated that in each million children in the United States, 32 will be diagnosed with ALL [20].

1.2 Link to the Environment

Although the causes of infant and childhood ALL are still under active scientific debate, a proposed mechanism involves intra-uterine translocations, caused by insufficient maternal folic acid levels (due to lack of ingestion), as well as by polymorphisms to genes involved in folate metabolism [2, 16]. This hints to the role of a finely-tuned relationship between genetic predispositions and inter-uterine environmental factors in the development of ALL [21].

It is now established that folate is essential for fetal health because its metabolism, which is catalyzed by methylenetetrahydrofolate reductase (MTHFR), generates or 'donates' molecules that contain one carbon atom, which are necessary for purine and pyrimidine synthesis in the cell. Purine and pyrimidine are the two types of subunits from which DNA and RNA are assembled, where they give rise to the bases of adenine and guanine, and cytosine, thymidine and uracil, respectively. In addition to DNA synthesis, purine and pyrimidine residues are also essential for the methylation of DNA, a process during which a methyl (CH_3) group is added onto the bases that constitute DNA. DNA methylation is an integral part of normal fetal development and growth [21]. MTHFR targets the metabolism products either into the DNA synthesis pathway or the DNA methylation pathway [21, 22]. When the levels of folate are low during pregnancy, there is a deficiency in the precursor that catalyzes the conversion of uracil into thymidine, and consequently, there is a lack in thymidine residues in DNA. This eventually leads to the development of chromosomal breaks and gaps that are translocation initiators (see section 1.3) [22]. In addition, this malfunction could lead to alterations in the normal methylation patterns of DNA at cancer-associated genes [21], such as Mixed Lineage Leukemia, thereby causing cancer (discussed in the section 1.3). Research has also been focused on polymorphisms of MTHFR that have better potential in the conversion of uracil into thymidine and, hence, have been associated with lower risks of cancer development. MTHFR also catalyzes the removal of

the methyl group on the amino acid methoinine to generate homocysteine; the reverse reaction is catalyzed by methionine synthase and is B_{12}-dependent [19]. When vitamin B_{12} is scarce, methionine, which as an essential amino acid, cannot be regenerated and employed in fundamental cellular processes.

While bioflavonoids represent a healthier food choice for adults, when ingested during pregnancy they may provide unfavorable growth conditions for the fetus. Bioflavonoids are inhibitors of an enzyme called topoisomerase II, which catalyzes the repair of double-stranded breaks in DNA. The presence of flavonoids in a mother's diet will therefore prevent the repair process from taking place [23] leaving DNA susceptible to translocations that lead to leukemia [18]. It is suggested that these events do not occur in adults because adult cells exhibit lower rates of proliferation compared to fetal cells [23].

1.3 The role of MLL and MCEF in ALL

The environmental conditions discussed previously are believed to implicate ALL in children by mediating a modification in a gene called *Mixed Lineage Leukemia* (*MLL*). The *MLL* gene is located on chromosome 11, at position q23, where q is the long arm of the chromosome (as opposed to p, the shorter arm); any rearrangements in which *MLL* is involved are hence referred to as 11q23 rearrangements. The *MLL* gene is repeatedly found rearranged in patients with ALL. The modifications in *MLL* occur in association with another gene, referred to as a partner gene; *MLL* is believed to have 64 partner genes, but over 104 different translocations [24-26]. The particular rearrangement that causes ALL is called a translocation, where one of the partner genes is physically moved to another location in the human genome, the *MLL* gene. Some of these rearrangements may also be further classified as reciprocal translocations, because both genes are moved, and thus substitute each others' locations in DNA. Typically, only a

portion of the gene is translocated, giving rise to fusion genes. During protein synthesis, these genes give rise to their respective proteins, producing fusion proteins (MLL-partner gene) that are implicated in ALL. Usually, only one of the alleles is involved in the translocation, and the production of fusion protein(s) is thought to cause disease either by gain of function or by interfering with the normal functioning of *MLL* [27].

This particular project focuses on one of these partner genes, *MCEF* (discussed in detail in section 1.4) [28, 29]. Although the insertion of *MCEF* with *MLL* is rare [29], the combined translocation frequency of *AFF1*, *AFF3* and *AFF4* with *MLL* represents the most common translocations that give rise to ALL [30-32]. These findings are the reason for the interest in the AFF family, particularly AFF4. The insertion of the *MCEF* gene with *MLL* produces an MLL-MCEF fusion protein whose N-terminus consists of *MLL* exons and C-terminus consists of *MCEF* exons (ins(5;11)(q31;q13q23)) (fig. 1) [29, 33]. It is important to note that the modification of *MLL* has also been proven to cause another form of leukemia called Acute Myeloid Leukemia (AML). In AML, it is the myeloblasts, rather than the lymphoblasts, that are abnormal and cannot differentiate into any of the cells that normally originate from myeloblasts: red blood cells, platelets and granulocytes (a type of white blood cells). It was previously estimated that AML accounts for approximately 5% and 16% of all infant and childhood cancers and leukemias, respectively, which is lower than the 30% and 80%, respectively, contributed by ALL [18, 34]. It is also worth noting that the frequency of translocations of *MLL* is estimated at 7-10% and 5% for ALL and AML, respectively [35]. AML also occurs *in utero* but no concrete evidence has yet been established on its association with folate and B_{12}.

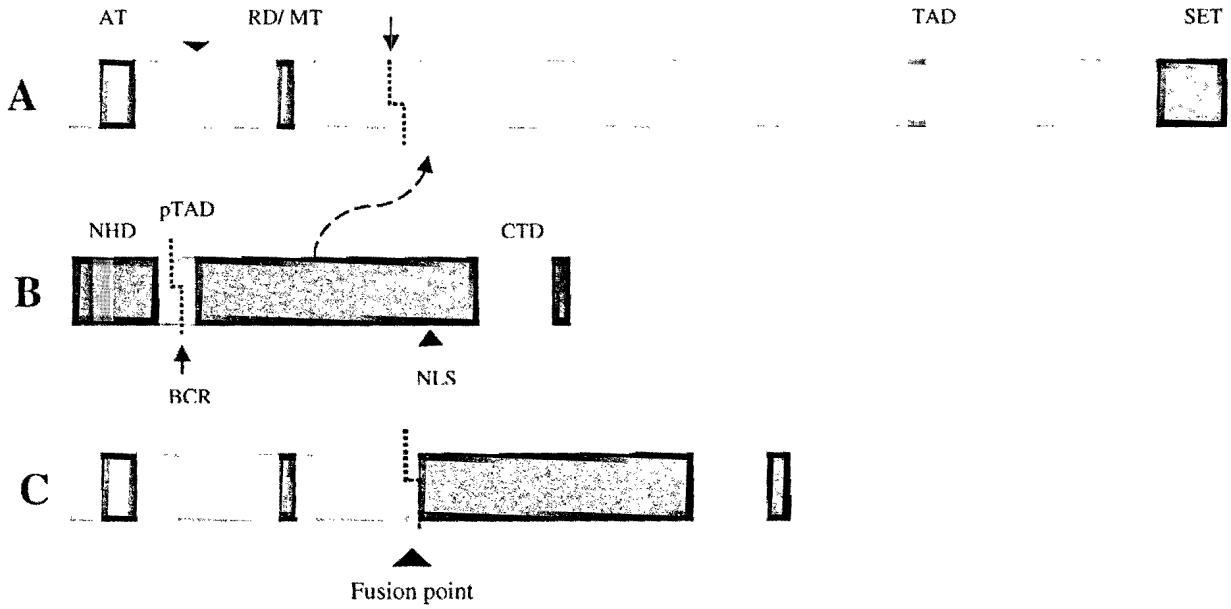


Figure 1. MLL-MCEF fusion protein. (A) Full-length *MLL*, with the AT hook region (AT), Nuclear Localization Signal (NLS), Repression Domain (RD)/ DNA methyltransferase homology domain (MT), breakpoint cluster region (BCR), Transactivation Domain (TAD) and ***Suppressor*** of variegation, Enhancer of zeste, Trithorax (SET) labelled. (B) Full-length *MCEF*. Labelled are N-terminal homology domain (NHD), putative transactivation domain (pTAD), breakpoint cluster region (BCR), nuclear localization region (NLS) and C-terminal homology domain (CTD). In (C), the fusion product of MLL and MCEF translocation is shown. The 5' region corresponds to the N-terminal of MLL, including the AT and RD/MT regions (exons 1- 10), and the 3' region corresponds to the portion of MCEF that is between amino acids 351- 1163, which includes part of the pTAD region, as well as the CHD (exons 6- 21).

1.4 Discovery of MCEF

The discovery of the positive transcription elongation factor b (P-TEFb) that regulates HIV-1 transcription elongation was first made in *Drosophila* [36, 37]; P-TEFb is one of the most important elongation transcription factors involved in the regulation of HIV-1 [38]. In humans, P-TEFb used to be referred to as Tat-transactivated-Kinase (TAK), until research revealed that TAK was actually the same protein as P-TEFb that was discovered in *Drosophila* [36]. P-TEFb is comprised of two subunits, namely a kinase called Cdk-9 and a kinase regulator called Cyclin-T_1 [39]. Several immunoprecipitation studies have also shown that Cdk-9 interacts with other cyclin T subunits, specifically cyclin T2A and cyclin T2B, which are both formed

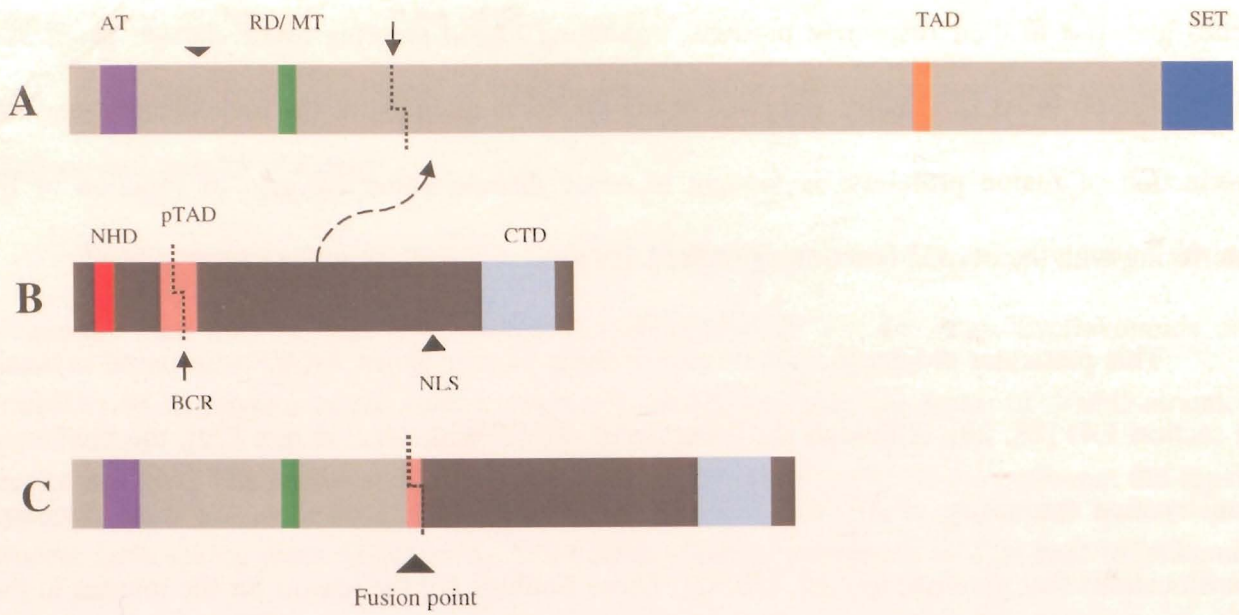


Figure 1. MLL-MCEF fusion protein. (A) Full-length *MLL*, with the AT hook region (AT), Nuclear Localization Signal (NLS), Repression Domain (RD)/ DNA methyltransferase homology domain (MT), breakpoint cluster region (BCR), Transactivation Domain (TAD) and ***Suppressor*** of variegation, Enhancer of zeste, Trithorax (SET) labelled. (B) Full-length *MCEF*. Labelled are N-terminal homology domain (NHD), putative transactivation domain (pTAD), breakpoint cluster region (BCR), nuclear localization region (NLS) and C-terminal homology domain (CTD). In (C), the fusion product of MLL and MCEF translocation is shown. The 5' region corresponds to the N-terminal of MLL, including the AT and RD/MT regions (exons 1- 10), and the 3' region corresponds to the portion of MCEF that is between amino acids 351- 1163, which includes part of the pTAD region, as well as the CHD (exons 6- 21).

1.4 Discovery of MCEF

The discovery of the positive transcription elongation factor b (P-TEFb) that regulates HIV-1 transcription elongation was first made in *Drosophila* [36, 37]; P-TEFb is one of the most important elongation transcription factors involved in the regulation of HIV-1 [38]. In humans, P-TEFb used to be referred to as Tat-transactivated-Kinase (TAK), until research revealed that TAK was actually the same protein as P-TEFb that was discovered in *Drosophila* [36]. P-TEFb is comprised of two subunits, namely a kinase called Cdk-9 and a kinase regulator called Cyclin-T_1 [39]. Several immunoprecipitation studies have also shown that Cdk-9 interacts with other cyclin T subunits, specifically cyclin T2A and cyclin T2B, which are both formed

from the same cyclin T gene that gives rise to cyclin T1, by alternative splicing. Once HIV-1 transcription has been initiated, a three-subunit complex that consists of Tat (a transactivator for HIV-1), P-TEFb and a growing mRNA stem-loop structure known as Transactivation Response (TAR) element, is formed. This particular arrangement causes the Cdk-9 unit of P-TEFb to hyperphosphorylate specific residues on an enzyme called RNA polymerase II (RNA pol II) that catalyzes the transcription of mRNA from DNA in eukaryotic cells [40-43]. RNA pol II consists of ten subunits that make up the catalytic core of the enzyme, and two additional subunits RNA polymerase II subunits 4 and 7 (RPB4/7) that are required for transcription initiation but that later dissociate from the core subunits. The largest of the core subunits, RNA polymerase II subunit 1 (RPB1), contains the Carboxy terminal domain (CTD) that is composed of repeats (26 in yeast, 52 in humans, 44 in fruit fly) of a conserved 7-amino acid sequence: Tyrosine-Serine-Proline-Threonine-Serine-Proline-Serine (YSPTSPS). It is suggested that the phosphorylation of CTD occurs on Serine 2 and Serine 5 residues of these repeats during transcription initiation and transcription elongation, and it is further demonstrated that these two residues are phosphorylated by two different kinase enzymes. Whereas the catalytic subunit of P-TEFb, Cdk-9, is responsible for phosphorylating Serine 2, Serine 5 is phosphorylated by another transcription factor named TFIIH. TFIIH is comprised of cyclin H (analogous to cyclin T1 of P-TEFb) and its' catalytic subunit Cdk-7, which phosphorylates Serine 5 [44]. More recently, Serine 7 was also identified as a site of CTD phosphorylation, where evidence suggests that Cdk-7 phosphorylates the residue prior to transcription initiation. The phosphorylation of the CTD of RNA pol II is necessary for elongation to commence at the long terminal repeat (LTR) region of the HIV-1 promoter. This is known as the tat-transactivation of HIV-1 transcription. It is suggested that phosphorylated CTD sequences amplify transcription not only by facilitating the

recruitment of mRNA processing (5'-end capping and 3'-end polyadenylation) enzymes to the elongating mRNA chain [45], but also by ensuring that the transcription is synchronized with crucial nuclear processes (through binding nuclear factors). It is important to note that only the cyclin T1 subunit (not cyclin T2A and cyclin T2B) of P-TEFb interacts with Cdk-9 to bring about the Tat-transactivation of HIV-1 [46].

The level of expression of P-TEFb is upregulated by the Ras/Mitogen-activated protein kinase (MAPK) pathway. Within the LTR of HIV-1, the core promoter is found. The HIV-1 core promoter is a stretch of transcription factor binding sites that extends from -200 to +50 from the site of transcription initiation; some of these binding sites include a TATA box, three SP-1 sites, two NF-κB enhancer motifs, and an RBEIII site. The RBEIII site is the Ras-responsive binding element III, and as the name suggests, is involved in a Ras/MAPK-signaling pathway. A transcription factor called RBF-2, which itself consists of a USF and TFIIi, binds RBEIII upon activation of the pathway [47-50].

The Ras/MAPK pathway is a signaling transduction pathway that regulates major molecular and metabolic processes in the cell (including cell proliferation, cell differentiation and cell death) with the aid of various proteins that function downstream of activation. In HIV-1, the T-cells Ras/MAPK pathway begins when an antigen is presented by an antigen-presenting cell (APC) and is recognized by the T-cell receptor (TCR), CD3, and its associated co-receptor (CD4 or CD8). Upon recognition, a protein known as Lck binds to the cytoplasmic portion of the CD4 co-receptor and phosphorylates and activates the CD3 receptor, allowing it to bind the cytoplasmic kinase Zap-70. Zap-70 phosphorylates and activates another protein, Linker of Activation of T cells (LAT), which in turn serves as a docking site for phospholipase Cγ (PLCγ), enabling it to catalyze the hydrolysis of Phosphatidylinositol phosphate and (PIP) into

diacylglycerol (DAG) and inositol Triphosphate (IP_3), which both serve as second messenger molecules in the cytoplasm. DAG production results in the activation of two pathways. The first pathway is the protein kinase Cθ (PKCθ) pathway, whose activation leads to the phosphorylation and activation of a mediator kinase known as I- κB (IKK). IKK is responsible for the consequent phosphorylation of the NF-κB inhibitor I- κB and its' dissociation from NF-κB (which can now be transported to the nucleus to bind the NF-κB motif in the HIV-1 core promoter). The second pathway activated by DAG is the Ras pathway, which is a cascade of kinase proteins that are subsequently activated to phosphorylate the downstream kinases in the pathway. DAG results in the activation of Ras-guanyl releasing protein 2 (Ras-GRP2), which exchanges Ras-bound GDP for GTP, thereby activating Ras; Ras then phosphorylates and activates Raf, which phosphorylates and activates mitogen-activated protein kinase/extracellular signal-regulated kinase (MEK). Next, MEK phosphorylates and activates extracellular signal-regulated kinase (ERK), which activates fos and jun (transcriptions factors of the activator protein, AP-1, family), as well as other proteins including TFIIi. These transcription factors are transported to the nucleus, where they bind to the HIV-1 LTR and drive cell proliferation. Meanwhile, IP_3 production leads to an increase in the influx of cytoplasmic Ca^{2+} ions, activating the protein calineurin, which dephosphorylates and activates Nuclear Factor of Activated T-cells (NFAT). This nuclear factor is then shuttled into the nucleus, where it causes the upregulation of cell proliferation and differentiation.

Importantly, Estable *et al.* [38] searched for P-TEFb-associated proteins, by immunoprecipitation and successfully co-immunoprecipitated a novel protein. This protein was called Major CDK9-interacting Elongation Factor-associated protein (MCEF), which also stands for the initials of its discoverer (Mario Clemente Estable Ferrero). After cloning its cDNA,

sequence analysis of the Open Reading Frame (ORF) of MCEF revealed an almost identical sequence to Af5q31 [38]. AF5q31 was the name given to a putative protein, which could be produced through translation of a theoretical ORF, found in a search for DNA translocation partners with MLL, in a case of ALL [28]. Because MCEF was the first designation of the actual protein and refers to the function of the protein, whereas AF5q31 is simply a chromosomal location, the MCEF designation for the protein is the correct one. However, the protein and gene are sometimes referred to in the literature as MCEF and/or Af5q31. More recently, the gene itself, coding for MCEF/AF5q31, has been designated as AF4 Family member 4 (AFF4) by the Human Genome Nomenclature Committee. Therefore, AFF4, MCEF and AF5q31 are all referring to the same gene, mRNA or protein. The designation AFF4 is in reference to being the fourth member of the AF4 family of proteins, which includes 3 other proteins; these are: AFF1/AF4, AFF2/FMR2 and AFF3/LAF-4. The *MCEF* gene codes for a protein that consists of 21 exons and 1,163 amino acids [51]. Experimentally, this protein has been found to have a molecular mass circa 140 kDa, although the theoretical calculated molecular mass of MCEF is 127 kDa [51]; the discrepancy could possibly be attributed to certain modifications that the protein undergoes after it is translated from mRNA (i.e. post-translational modifications). The difference could also be the result of multiple isoforms of the same protein being expressed in the cell.

One form of modification is phosphorylation/dephosphorylation. The addition of a phosphoryl group to a protein residue can change its conformation and thus potentially its function in the cell. This process normally happens in response to certain cellular changes carried through signal transduction pathways [52]. The presence of a free hydroxyl group in serine, tyrosine and threonine residues renders these amino acids the most susceptible to kinase and

phosphatase activity. Kinases are normally categorized in terms of the residues they phosphorylate (e.g. serine-specific kinases), where each group of kinases recognizes and phosphorylates the same residues, a consensus sequence, in a protein. For example, as mentioned previously, the phosphorylation consensus sequence that CDK9 and CDK7 recognize in the CTD

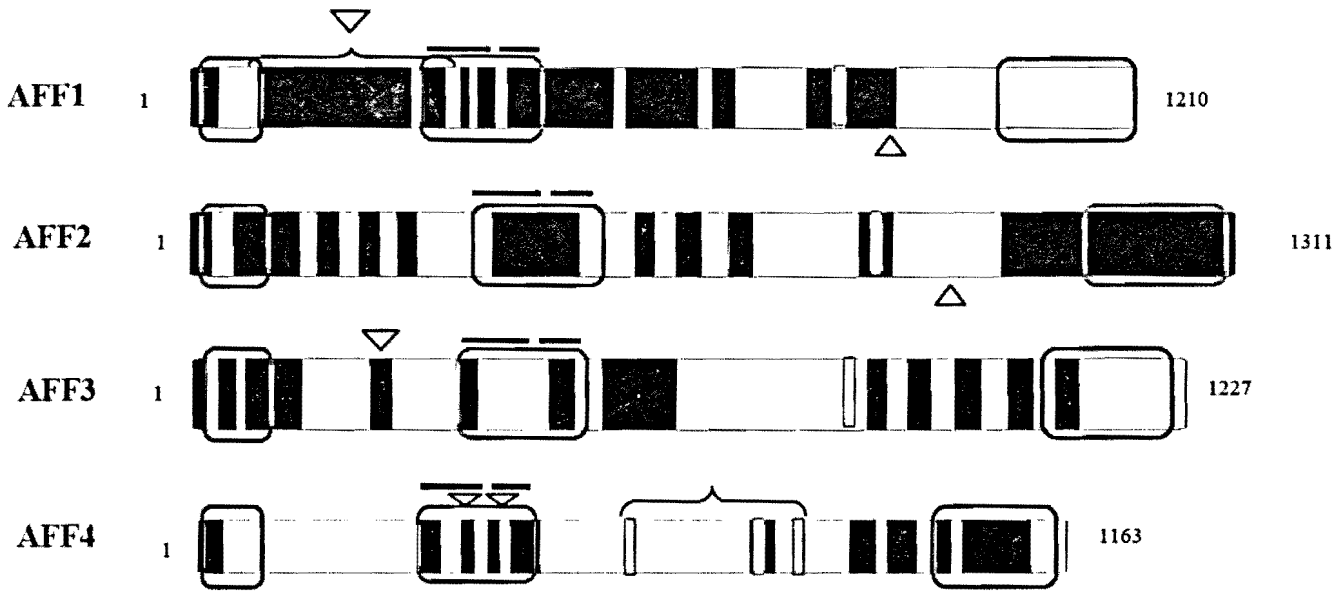


Figure 2. Human AF4 family members. The 4 members of the AF4 family of proteins, AFF1/AF4, AFF2/FMR2, AFF3/LAF-4 and AFF4/AF5q31/MCEF, are shown, along with their homology domains. The numbers at the N and C terminal ends of the protein indicate the first and last amino acids of the full-length protein; alternating white and grey boxes indicate consecutive exons; grey boxes indicate nuclear localization signals (NLS)/NLR; blue boxes indicate NHD; green boxes indicate pTAD, red boxes indicate CHD; line segments indicate Af4-Laf4-Fmr2 homology region (ALF) and poly serine region (pSer); black triangles indicate GTP binding sequence (GTP) for AFF1 and helix-loop-helix dimerization signature (HLH) sequence for AFF2/FMR2; inverted black triangles for AFF1, AFF3, AFF4 indicate breakpoint regions for fusion with MLL.

is YSPTSPS, where each kinase phosphorylates non-overlapping residues. Several studies have reported that MCEF is post-translationally modified, where at least 23 different phosphorylated peptides were found, particularly in the poly Serine domain of the protein [53-57]. Research groups were collectively able to identify 44 different residues on MCEF that were subject to phosphorylation; 39 of these were serine residues, 3 were tyrosine residues and 2 were threonine

phosphatase activity. Kinases are normally categorized in terms of the residues they phosphorylate (e.g. serine-specific kinases), where each group of kinases recognizes and phosphorylates the same residues, a consensus sequence, in a protein. For example, as mentioned previously, the phosphorylation consensus sequence that CDK9 and CDK7 recognize in the CTD

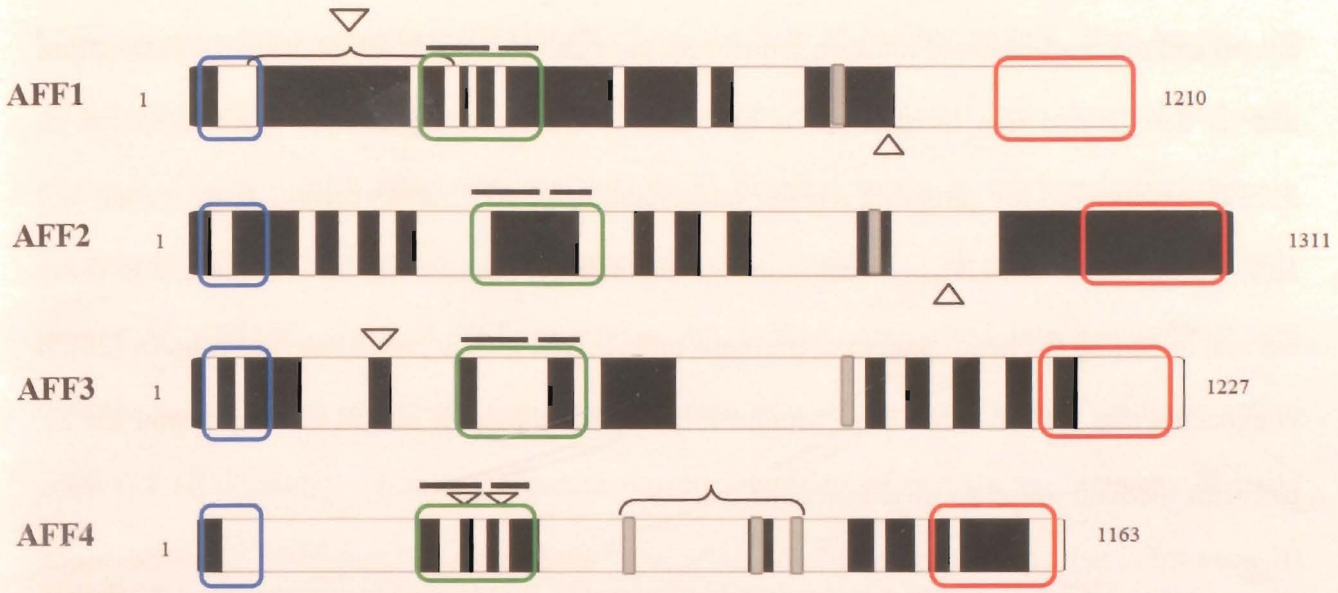


Figure 2. Human AF4 family members. The 4 members of the AF4 family of proteins, AFF1/AF4, AFF2/FMR2, AFF3/LAF-4 and AFF4/AF5q31/MCEF, are shown, along with their homology domains. The numbers at the N and C terminal ends of the protein indicate the first and last amino acids of the full-length protein; alternating white and grey boxes indicate consecutive exons; grey boxes indicate nuclear localization signals (NLS)/NLR; blue boxes indicate NHD; green boxes indicate pTAD, red boxes indicate CHD; line segments indicate Af4-Laf4-Fmr2 homology region (ALF) and poly serine region (pSer); black triangles indicate GTP binding sequence (GTP) for AFF1 and helix-loop-helix dimerization signature (HLH) sequence for AFF2/FMR2; inverted black triangles for AFF1, AFF3, AFF4 indicate breakpoint regions for fusion with MLL.

is YSPTSPS, where each kinase phosphorylates non-overlapping residues. Several studies have reported that MCEF is post-translationally modified, where at least 23 different phosphorylated peptides were found, particularly in the poly Serine domain of the protein [53-57]. Research groups were collectively able to identify 44 different residues on MCEF that were subject to phosphorylation; 39 of these were serine residues, 3 were tyrosine residues and 2 were threonine

residues (see table A-1). These residues are spread throughout the length of the protein and are not clustered within the serine rich region (fig. 2), and because most of the residues are serine residues, it would be logical to assume that serine-specific kinases would be responsible for the phosphorylation of MCEF.

Although MCEF is implicated in multiple diseases, and although the cDNA has been cloned and the chromosomal location sequenced, the exact role of MCEF in normal cell function, and its deregulated role in diseases, is not known. Therefore, the further characterization of MCEF is warranted for garnering a better understanding of ALL, azoospermia, body weight and HIV-1 replication, as well as normal function. Importantly, because MCEF is expressed in many tissues, including the heart, pancreas, placenta, and skeletal muscle, and even in the brain [28], it is expected that MCEF either has a similar role to play in most if not all tissues, or that MCEF isoforms mediate tissue-specific functions.

Since MCEF is part of a larger family of proteins, it is logical to assume that MCEF then shares some functional and structural similarities with the other proteins in the AF4 family. The AF4 proteins are a family of putative transcription factors with hypothetical conserved transactivation domains (see fig. 2) [58-61], although the actual existence of a transactivation domain in MCEF flies against its repressive function [38, 51]. The homology domains that these proteins share are: N-terminal homology domain (NHD), putative transactivation domain (pTAD), Serine-rich region (pSer), AF4-LAF4-FMR2 region (ALF), and Carboxy-terminal homology domain (CHD) (fig.2). All 4 members are localized to the nucleus of the cell, and therefore also contain nuclear localization signals (NLS).

Nuclear localization signals are sequences within a protein that direct the transport of that

protein to the nucleus. NLS are usually composed of positively charged amino acids (e.g. lysine, arginine), which enable proteins to be recognized by importin α, a protein that, upon binding the NLS, binds a second protein, importin β. Together, importin α and importin β form a 3-subunit complex with the NLS-containing protein which targets the bound protein to the nuclear pore complex (NCP) and aids its transport across the nuclear envelope. In the nucleus, the importin-bound protein dissociates from the complex with the help of another protein, Ran, leaving the NLS-bound protein free to bind DNA. NLS sequences exist in two types: classical and non-classical. Classical NLS sequences are thought to function solely as nuclear import signals, whereas non-classical sequences may sometimes also signal nuclear export rather than only import. In addition, classical NSL sequences are further categorized into monopartite and bipartite sequences. A monopartite sequence consists of a single short "stretch" of basic amino acids (~7-10 residues), whereas a bipartite sequence consists of two shorter stretches of basic amino acids (~2-5 residues) separated by a "spacer" of non-basic amino acids that is between 10 -12 residues long [62]. Non-classical sequences, on the other hand, are not dominated by basic amino acids stretches; rather, they are composed of hydrophobic residues. The most well-known example of non-classical NLS sequences is the 38-residue M9 sequence found in the human mRNA-binding protein (hnRNP) [62].

The numbers, positions and types of the nuclear localization signals on the proteins vary (fig.2). For example, the C-terminal of AFF1/AF4 contains 6 putative classic sequences, two of which overlap to form a seventh, bipartite sequence [63], whereas AFF2/FMR2 and AFF3/LAF-4 contain 2 and 1 classic putative bipartite sequences, respectively [58, 61]. MCEF contains 7 nuclear localization signals, all contained within a nuclear localization region that spans 293 amino acids [27]. Research performed by Dr. Estable's team demonstrated that of these 7

signals, all are classical sequences (located in exons 11, 12, 13 and 14) with the exception of one, which has a non-classical sequence (located in exon 13) [51]. Deletion analysis suggested that only 3 of these signals are functional and able to target MCEF to the nucleus: 2 bipartite sequences, named NLS 1 and NLS 5 in exons 11 and 12, respectively, and the non-classical sequence in exon 13. The remaining NLS were not able to label the protein for nuclear transport without the presence of NLS 1 and NLS 5 [51]. As of yet, it is unclear why only 3 NLS are functional and which amino acids are essential for functionality of an NLS and nuclear localization of MCEF. Further research, specifically site-directed mutagenesis experiments, is required for gaining a better understanding of the NLS of MCEF.

AFF1/AF4 is a protein encoded by the AFF1/AF4 gene that is located on chromosome 4 in the genome and is 1,210 amino acids long (~ MW 131 kDa). Research has shown that the level of AFF1/AF4 expression is high in placenta, human heart and skeletal muscle [28, 31]. As was first shown for AFF4/MCEF/Af5q31 by Estable *et al.*, AFF1 is thought to form a complex with P-TEFb (through direct interaction) and other transcription factors to bring about conformational changes in DNA that are crucial for transcription elongation [64]. Mutations in AFF1/AF4 are associated with cerebellar ataxia, a disease that is characterized by a patient's loss of muscle control that leads to the lack of coordination (ataxia) in many part of the body, including the eyes and limbs. This is thought to occur due to the accumulation of mutated AFF1/AF4 in the cells of the cerebellum (the part of the brain that controls muscle coordination); mutations in the protein compromise the ability of degradative enzymes, Siah-1 and Siah-2, to recognize and degrade it [65, 66]. Siah-1 and Siah-2 belong to the seven in absentia homolog (SIAH) family of enzymes that mediate the degradation of proteins not only by carrying out the process itself, but also by labeling the target proteins for degradation. More relevant to this

paper, AFF1/AF4 is recognized for being the most frequent fusion partner of MLL, where a significant fraction of infant and childhood ALL cases are caused by the t(4;11)(q21;q23). In fact, this protein was initially discovered as an MLL-fusion partner [67-69].

AFF2/FMR2 is the second member of the AF4 family. The gene for this protein is located on chromosome X and is translated into a 1,311 amino acid protein (~MW 141 kDa). It is the longest of the AFF members, and as of 2010, no reports were published on its implication in ALL and is therefore the only member that is not associated with MLL translocations or leukemia. AFF2, however, is the primary culprit that causes familial mild mental impairment. Fragile X E (FRAXE) is a form of mental impairment that is correlated with concentration, communication and learning difficulties, as well as autistic characteristics in patients [70]. Research has established that FRAXE is linked to a fragile site, E, on chromosome X (Xq28); the fragile X E (FRAXE) site belongs to the folate-sensitive fragile family that have been repeatedly associated with mental retardation [71]. These sites are fragile because they are prone to undergo modifications in the adjacent Cytosine-phosphate-Guanine (CpG) islands, which is in turn due to the presence of breaks and gaps in these chromosomal regions that arise when DNA synthesis is temporarily arrested. CpG islands are stretches of DNA that contain a high number of CpG sites, that is, positions where a cytosine occurs next to guanine residue, linked together by a phosphate molecule. When CpG islands are located nearby genes that are normally expressed in the cell, the CpG residues are unmethylated. Moreover, under normal circumstances, the fragile site exhibits between 6-35 CCG copies; in this manner, nearby genes are expressed. However, in mild mental impairment, there are over 200 copies of the CCG repeats, and the CpG island is 100% methylated [72, 73]. These modifications cause the fragile site to occupy a larger region on the chromosome and inhibit, or silence, the expression of associated genes. In the case of

FRAXE, the gene that is silenced is fragile X mental retardation 2 (FMR2), or AFF2 [58, 59, 71]. Research in this field led to the discovery of protein called Ox19, whose gene overlaps that of AFF2 on chromosome X and whose sequence is identical to AFF2, but is significantly truncated and processed at the 3' end [74]. Ox19 is therefore considered an isoform of AFF2 that is produced by alternative splicing [58]. AFF2 is highly expressed in brain tissue and acts as a transcription activator [75].

The AFF3/LAF4 gene is located on chromosome 2 and it codes for a protein 1,277 amino acids long (~ MW 134 kDa). As is the case with AFF1/AF4 and AFF4/MCEF/Af5q31, AFF3/LAF4 was initially discovered as a fusion partner of the MLL gene [60]. This protein, however, causes ALL through 3 different fusions with MLL: 2 translocations (t(2;11)(p15;p14) [76] and t(2;11)(q11;q23)) [77] and an insertion (ins(11;2)(q23;q11.2q11.2)) [78]. More recently, the level of AFF3/LAF4 expression was found to be higher in over 20% of breast cancer cases, suggesting that it could potentially be a proto-oncogene [79]. A proto-oncogene is a gene that under normal circumstances performs usual cellular activities, but, when mutated, is converted into its' dysfunctional form, an ocnogene, where it can result in uncontrolled cell proliferation, i.e. cancer.

In addition, a microdeletion of 500 kb on chromosome 2 was observed in some cases of Nievergelt syndrome (NS) [80]. This condition is characterized by abnormal development and deformation of the bones in the lower regions of arm and leg, specifically the radius, ulna, tibia and fibula bones [80]. Chromosomal analysis of the microdeletion revealed that it encompasses the gene that codes for AFF3/LAF4, implying that AFF3/LAF4 could partially cause the development of NS [80]. A recent study has also proposed the possibility that AFF3/LAF4 may be associated with rheumatoid arthritis (RA) [81]. According to Barton *et al.*, there are 13

potential genes, including AFF3/LAF4, that may contribute to RA susceptibility, as well as susceptibility to other autoimmune diseases (e.g. multiple sclerosis and type 1 diabetes), suggestive by the occurrence of variants of these genes in the majority of disease cases. AFF3/LAF4 is the newest autoimmune susceptibility gene identified by the group [81], suggesting that it may also be implicated in autoimmune diseases, as well as leukemia. In normal cells and tissues, AFF3/LAF4 is expressed in the placenta, adult brain and heart, as well as fetal brain, liver and kidney [77].

As mentioned previously, fragile sites are a product of non-arbitrary breaks in specific regions in DNA that occur during DNA synthesis in metaphase [82], and AFF4/MCEF/Af5q31 is implicated in ALL through its association with MLL. The breakpoints of both *MLL* and *MCEF* have been recently mapped to predisposed chromosomal "fragile sites" in DNA [83]. In fact, this is true for over half of the gene sets (52%) responsible for inflicting ALL in infants, where either one (40%) or both (12%) of the genes are mapped to fragile sites. Additionally, over 65% of leukemic translocation breakpoints are actually common fragile sites that exist in all individuals, which means that at any one point, translocations may be easily induced, a fact that again reinforces the role of environmental factors and chemical agents in the development of disease [83]. The existence of fragile sites in *MLL* and its partner genes suggests that the base pair at which the chromosome breaks may not always be identical for the same gene, rather, specific regions exist where the probability of chromosomal breaks is significantly higher: the breakpoint cluster region (BCR) [84]. For example, the *MLL* BCR is located between exons 5 and 11 of the genes [84]. Thus, the breakpoints of *MLL* with *MCEF* that have been so far documented are specific to the insertion/translocation that occurs; in the insertion documented by Taki *et al.* in 1999, the *MLL* breakpoint is between exons 8- 10, whereas in the translocation reported by

Imamura *et al.* in 2006, the breakpoint is determined to be in exon 6 [28, 33].

Interestingly, a very recent study has shown that AFF4 is actually a shared subunit of many of the other MLL chimeric proteins associated with leukemogenesis [85]. The MLL chimeras tested were MLL-ELL1, MLL-ENL, MLL-AFF1 and MLL-AF9, and once these were purified and analyzed, AFF4 was found in association with all 4 fusion proteins, as first shown by Estable *et al.,* [38]. Moreover, the group discovered complexes that contain elongation factors ELL1, ELL2, ELL3 and P-TEFb (detected were subunits cyclin T_1 and Cdk9) with AFF4 (as a central component), dubbed super elongation complexes (SEC), an updated form of the association discovered previously [38, 85]. These complexes were able to phosphorylate the CTD of RNA polymerase II, and their stability was dependent on AFF4. The results seem to suggest that not only can MCEF inflict ALL through its direct interaction with MLL, but it can also influence the interaction of other fusion partners with MLL via SEC, making it an important regulator in the development of infant leukemia and a possible target for its treatment; this implies that ALL is a disease of transcription elongation [85].

MCEF also interacts with Cdk9, the catalytic subunit of P-TEFb. In fact, a previous study conducted by my faculty supervisor and his team suggested that MCEF had the ability to down-regulate HIV-1 transcription and repress its replication. MCEF repressed HIV-1 transactivation by 60% in the presence of Tat [38]. MCEF is thought to repress HIV-1 transcription and replication only in the presence of Tat; experiments performed in the absence of Tat showed that the MCEF did not significantly repress HIV-1 transcription [51]. Taking this data together with the fact that MCEF interacts with P-TEFb, one possibility is that in order to down-regulate HIV-1 replication, MCEF must interfere with the formation of the ternary complex necessary for transcription elongation [51]. Also in agreement with the role of MCEF as

a repressor, when tethered to a T7 promoter in HeLa cells, MCEF was not able to transactivate transcription, in contrast to what occurs with AFF1/AF4, AFF2/FMR2 and AFF3/LAF4 [75]. Together, these findings imply that MCEF is in fact a negative HIV-1 transcription regulator, making it a potential drug or therapeutic target for the treatment of HIV-1/AIDS.

In addition, a study in mice demonstrated that MCEF is required for normal spermatogenesis, since MCEF-knockout mice exhibited azoospermia [86]. Knock-out mice are subjects that have only one wild type (functional) copy of the gene, and are therefore lacking the other allele of the MCEF gene, i.e. they are "MCEF-deficient". The study demonstrated that the levels of Af5q31 expression in the testes of adult mice are higher than its expression in other tissues (heart, liver, and kidney). Moreover, 87% of Af5q31-deficient mice died either neonatally or *in utero*, while the remaining 13% survived as wild type mice, only with a completely compromised ability to reproduce. Analysis revealed that infertility was brought about by the lack of mature spermatozoa, and that the absence of a functional Af5q31 protein affected the integrity of testicular somatic cells, despite the normal expression of some of the other important proteins required for sperm production. Collectively, these results propose a new area for the transcriptional regulatory activity of MCEF in embryonic development and germ cell differentiation and survival [86].

The overall purpose of this project was to contribute towards the characterization of the MCEF protein in terms of its structure and function. To achieve this, several approaches were taken. These are: (i) I produced an updated phylogenetic analysis to show the evolutionary relationships among the AFF proteins; (ii) based on the alignments of AFF1 to AFF4, I designed MCEF-specific peptides and had these injected into rabbits, to create MCEF-specific anti-MCEF antisera, as a tool to characterize MCEF; (iii) using the rabbit-MCEF antisera as a tool, I show

some western blot experiments that suggest the existence of MCEF protein isoforms, potentially explained by multiple splicing; (iv) I show some preliminary immunomagnetic separation of full-length MCEF from nuclear extracts of HeLa cells; (v) I show some experiments suggesting phosphorylated isoforms of MCEF; (vi) I show deletion analysis experiments to examine the effect of MCEF on the tat-activated transcription of HIV-1; (vii) I show some site-directed mutagenesis experiments on HIV-1 LTR (specifically RBEIII) that severely impair the ability of HIV-1 to replicate; (vii) I discuss the labs goal of tethering MCEF to RBEIII, as a future therapy against AIDS.

2. MATERIALS AND METHODS

2.1 Phylogenetic Analysis

The program used to carry out the phylogenetic analysis was MacVector (11.0.2), and the sequences used were curated by me, directly from NCBI (GenBank). Complete sets for the 4 members of the AF4 family were found for the human, chimpanzee, monkey, rat, mouse and horse species. The following is a list of the accession numbers of the sequences used for each protein; **AFF1**/AF4: P51825 (*Homo sapiens*), 114594998 (*Pan troglodytes*), XP_001095779 (*Macaca mulatta*), NP_001100676 (*Rattus norvegicus*), AAU93698 (*Mus musculus*), XP_001916121 (*Equus caballus*); **AFF2**/FMR2: NP_002016 (*Homo sapiens*), NP_001009042 (*Pan troglodytes*), BAE02338 (*Macaca mulatta*), XP_001054673 (*Rattus norvegicus*), NP_032058 (*Mus musculus*), XP_001914921 (*Equus caballus*); **AFF3**/LAF4: AAH36895 (*Homo sapiens*), XP_525831 (*Pan troglodytes*), XP_001104149 (*Macaca mulatta*), XP_343560 (*Rattus norvegicus*), NP_034808 (*Mus musculus*), XP_001492374 (*Equus caballus*); **AFF4**/AF5q31/MCEF: NP_055238 (*Homo sapiens* AFF4), AAF18981 (*Homo sapiens* AF5q31), AAM00184 (*Homo sapiens* MCEF), XP_517928.2 (*Pan troglodytes*), BAE00514 (*Macaca mulatta*), NP_001100471 (*Rattus norvegicus*), NP_291043 (*Mus musculus*), XP_001504471 (*Equus caballus*). Three nearly identical sequences for human member 4 were used, one corresponding to each of the following: AFF4 (hypothetical), AF5q31 (generated by research) and MCEF (generated by research). The previous sequences were all aligned using ClustalW in MacVector, using the default settings. The phylogenetic analysis option was used to first generate the best tree by Neighbour Joining, and the bootstrap option was then used to generate the node values by running 32,000 bootstraps. The sequence of the Lilliputian protein in Fly *Drosophila* (accession no. NP_722863) was used as the outgroup in the generation of both the neighbour joining and bootstrap trees. Lilliputian is a protein in fruit fly that is crucial for normal

development and growth. It shares sequence and domain homology with the four members of the AFF family, and is therefore the only AFF member in fruit fly. This suggests that Lilliputian and the AFF members share a common ancestor [87], making Lilliputian a suitable outgroup to perform this analysis.

2.2 SDS-PAGE and Western Blot Analysis

2.2.1 SDS-PAGE and Western Blots

Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE) was used to separate proteins from whole cell and/or nuclear extract samples according to their molecular weights. The concept of separating proteins using polyacrylamide and electrical current, i.e. PAGE, has been used since 1930, but the technique was refined by Laemmli in 1970 to include two gels, stacking and separating, and SDS to improve the efficiency of protein separation [88]. Laemmli's contribution to this field was so significant, that the buffer he developed is now referred to as the Laemmli buffer and used extensively in research.

The separating/resolving gel was prepared to 10% polyacrylamide, by mixing 2.5 mL 30% acrylamide, 5 mL deionized water, 2.5 mL SDS Tris-HCl lower buffer pH 8.8, 73.05 µL 10% ammonium persulfate (APS) and 21.9 µL Tetramethylethylenediamine (TEMED). Polyacrylamide is a polymer composed of acrylamide subunits, and in the context of SDS-PAGE, can absorb water and form a gel containing channels, or pores, through which proteins can pass (according to their size) [89]. APS is a polymerization initiator and TEMED is a polymerization co-initiator or activator, and they are used in combination to bring about the polymerization of the gel [89].

The stacking gel was prepared by mixing 0.75 mL 30% acrylamide, 3 mL deionized water, 1.25

mL SDS Tris-HCl upper buffer pH 6.8, 29.25 μL 10% APS and 19.5 μL TEMED. Depending on the volume of sample, the appropriate volume of sample buffer loading dye (mixing 12% SDS, 40% glycerol, 30% β-mercaptoethanol, 300 mM DTT (the two latter reagents disrupt disulfide bridges to ensure denaturation of the peptides), and 120 mM EDTA (chelates Ca^{2+} and Mg^{2+}, cofactors that are required by degradative enzymes, thereby preventing degradation of the proteins), 1 mg/mL bromophenol blue (a dye that makes it possible to visualize and track the progress of migration), 375 mM Tris-HCl pH 6.8 makes a 6x stock solution) was added and the samples were allowed to denature for 3 minutes at 95°C (activates and accelerates the denaturation of the proteins by SDS, β-mercaptoethanol and DTT) . The samples were put on ice, spun down in a centrifuge at 10,000 RPM then loaded onto the gel (at room temperature), where they were run at 130 mV until the dye front ran off the gel (approximately 45 minutes). Note that 4 different cellular extracts were used: HeLa nuclear extract NE13 (prepared in lab, [38]), Santa Cruz HeLa nuclear extract (SCNE), Santa Cruz HeLa whole cell lysate (SCWCE) and Santa Cruz HeLa whole cell lysate + alkaline phosphatase (SCWCE+AP) (the latter three were purchased from Santa Cruz Biotechnology, Inc., order nos. sc-2120 , sc-2200, sc-24677, respectively) .The gel was transferred to a western blot apparatus (Idea Scientific, product # 4015). The proteins were allowed to transfer for 60 minutes at 12 V. The membrane was removed and left to block in 5% skim milk in Phosphate Buffered Saline (PBS)-Tween 0.01% (PBS-T) for 30 minutes, and subsequently washed in PBS-T for 5 minutes. The membrane was incubated with primary antibody for 30 minutes. Several MCEF primary antisera were used in different experiments; the name of the antisera is provided in italics, followed by the amino acid numbers of the MCEF protein against which each was generated in brackets: *827* (1- 420), *826* (1- 530), *823* (1- 715), *AFF4* (189- 206), *N36* (857- 870), *N37* (1139- 1154) (all generated from

rabbits). All antisera except N36 and N37 were generated by Dr. Estable (827, 826, 823) and his colleague Dr. J. Brissette (Harvard-antisera designated as AFF4) in research labs; N36 and N37 were purchased commercially from BioSynthesis Inc. The antisera used in these experiments were all diluted to 1:1,000 in PBS-T, except N36 and N37, which were diluted to 1:500 instead due to the weaker signal detected by these two antibodies during the preliminary tests. The bleeds used were different for each antibody; this was determined by preliminary testing of each bleed of each primary antibody to identify the sample that provides the strongest signal: 16 weeks 823, 16 weeks 826, 22.5 week 827, 8 weeks N36, 8 weeks N37. Once incubation with primary antibody was complete, the membrane was washed 3 times with PBS-T, for 5 minutes each. The membrane was incubated with secondary antibody (Pierce, product # 32460) targeted aginast the primary antibody. The secondary antibody used in the experiments was anti-rabbit to make the recognition and binding of the membrane-bound primary antibody possible; antibody recognition is species-specific and the primary and secondary antibodies must therefore be obtained from the same species (in this case, rabbit). The secondary antibody was diluted to 1:20,000 (to eliminate non-specific binding of the secondary antibody to the primary antibody). The dilution was performed in PBS-T and the membrane was incubated for 30 minutes. Next, the membrane was washed twice for 5 minutes each, then twice for 15 minutes each. To visualize bands on the membrane, it was soaked in a chemiluminescent solution (Pierce, product # 34080) for 5 minutes. The chemiluminescent solution is generated by mixing two individual component solutions in equal parts: enhancer and hydrogen peroxide. The secondary antibody is linked to a reporter enzyme (the enzyme linked in the secondary antibody used here is horseradish peroxidase (HRP)) that oxidizes the enhancer component into a product that is oxidized again by the peroxide component. The final product of this reaction emits a light (chemiluminescent)

signal at the regions on the membrane where protein is detected; the intensity of the signal is proportional to the amount of detected protein. The membrane is then exposed to an autoradiography film and developed in the dark using a film processor (Konica Medical, Model SRX-101A). All incubations and washes were performed at room temperature on a gentle horizontally-rotating shaker.

2.2.2 Development of antisera

As mentioned previously, the antisera used in the western blot experiments were obtained from different sources. Whereas the 82 series and AFF4 were developed by Dr. Estable and colleagues, the N36 and N37 antisera were developed commercially by Biosynthesis, Inc; the procedure that gave rise to the all antisera, however, is the same. The peptide that is selected for designing the antibody must contain antigenic sites, i.e. hydrophilic residues that will most likely be located on the outer regions of the protein and therefore enable the antibody to recognize the target protein. When designing the N36 and N37 antisera, the antigenic sites were identified using a web-based program, Gizmo, recommended by Biosynthesis, Inc. The sequence of each exon was fed into the program and the output would provide different peptides (15-18 residues long), each with a unique antigenic value; the peptide with the highest antigenic value is selected. The sequences selected are then synthesized chemically to generate antigens that target the protein of interest (MCEF). Synthesis here is carried out starting at the C-terminal and ending at the N-terminal (opposite to what occurs during protein synthesis in the cell), where the carboxyl group of one residue is linked to the amino group of the next. The antigens are mixed with an adjuvant, an agent that boosts the immune response of the recipient of the antigen. This mixture is then injected into the recipient (rabbit) in order to develop antisera against the injected antigen. To assess the immune response and to extract the antibodies, blood is collected at

several time points (1 week, 3 weeks, 8 weeks, 12 weeks, final bleed) following injection. A test bleed is usually collected before injection; this bleed is used as a control to ensure that the antibodies were absent pre-injection and that immune response was elicited by the antigen. The bleeds are tested for cross-reactivity by ELISA experiments. Once the antisera are to be used in the laboratory, each is diluted in buffer to ensure that non-specific binding does not occur.

2.3 Phosphorylation Experiments

2.3.1 Immunomagnetic Separation

The hands-on work was done by Beverlee Buzon, as follows. The magnetic beads used in this experiment, Dynabeads Protein G, were purchased from Invitrogen (Cat. # 100-07D). The vial was shaken vigorously to suspend the beads then 50 µL of Dynabeads was aliquoted (with a tip cut at its end) into a 1.5 mL Eppendorf tube, and subsequently washed with citrate-phosphate buffer with 0.01% Tween-20. Tween-20 is a polymer composed of oxyeythlene subunits, where 20 refers to the number of monomers (i.e. the length of the chain). It is a surfactant that is used here as an emulsifier that prevents non-specific binding. The buffer is made by combining 0.47 g anhydrous citrate with 0.92 g bibasic sodium phosphate in 100 mL water. 200 µL of 827 antisera (bleed 22.5 weeks) (see section 2.2) and 100 µL citrate-phosphate buffer were then added to the tube and shaken gently on a gyrator, in a glass dish at 100 rpm, for 40 minutes at room temperature. Following, citrate-phosphate buffer with 0.01% Tween was first used to wash the coupled beads, and a second time with 0.2M ethanolamine with a Dynabead-specific magnet (supplied with the beads) to separate the beads from solution. Dimethyl pimelimidate (DMP) is a cross-linking agent (imidoester) that was used to cross-link the antibody to the G-protein. This was prepared by dissolving 5.4 µg in 1 mL 0.2M ethanolamine pH 8, an organic solvent and weak base that will not cross-react with DMP. The cross-linking was performed by incubating in

the Eppendorf tube for 30 minutes. The reaction of cross-linking the primary amine of the antibody and the G protein was stopped by incubating in 50 mM Tris-HCl pH 7.8 for 15 minutes. The beads were rinsed three times using PBS-T 0.01% and a magnet. Once the beads were cross-linked, 200 µL of HeLa S3 nuclear extract and 100 µL of PBS-T were added to the Eppendorf tube. The HeLa S3 (a subclone grown in suspension) nuclear extracts used were NE13, prepared by Dr. Estable [38]. Upon attaching the tube to the second hand of a clock to allow for even rotation, it was incubated for 1 hour at 4°C. The beads were washed with PBS-T thrice with the use of a magnet. Finally, to elute the protein from the antibody, 20 µL of 2 M citrate pH 2.0 were used twice and the sample was added to 20 uL 1M Tris-HCl, pH 7.8. The elution was dependent on citrate because an acid is required to change the charge, and therefore the confirmation, of the protein to bring about unbinding.

2.3.2 Alkaline Phosphatase Assay

The hands-on work was done by Beverlee Buzon, as follows. The calf intestinal phosphatase (CIP) enzyme used for the reaction was purchased from the University of Toronto Medstore (product # M0290S). To dilute the glycerol to 4% of the total volume, 1.6 units of CIP/µg nuclear extract with dilution buffer were used. The reaction was incubated for 30 minutes at 37°C and stopped by adding sample buffer loading dye. The loading dye is the same dye used in the western blot experiments and it contains SDS and β-mercaptoethanol, to achieve complete protein denaturation. The samples were later either used in an SDS-PAGE or an immunomagnetic separation.

2.4 MCEF Deletion Analysis

The hands-on work was done by Robert Hopewell, as follows. Four MCEF constructs were generated; each only coding for partial forward MCEF sequences (amino acids 1-715, 402- 745

and 694- 1163) or the full reverse MCEF sequence (1163- 1). The MCEF Open Reading Frame (ORF) was amplified using the forward primers with Xho1 and the reverse primer with BamHI. The primers used for the deletion constructs are given below. PCR, cloning and sequencing steps were performed on these constructs using standard techniques.

1- 715 F: (5'-CCGCTCGAGCGGATGAACCGTGAAGACCGG-3')

R: (5'-CGCGGATCCTGTCTGCTGGAGGGCATGC-3')

402- 745 F: (5'-GCCGAGGAGTACACCAGGAAGTAACTCTG-3')

R: (5'-CGCGGATCCTGGCACATTTTTCTTTTCCCCCTTGG-3')

694- 1163 F: (5'-CCGCTCGAGATGGAAGAGAAGGAACTTC-3')

R: (5'-CGCGGATCCCTAGATATCAACTTGGCATCCTG-3')

The HIV-1 LTR was cloned into a luciferase reporter construct, pRL-CMV (Promega, product # E2261), containing the *Renilla* luciferase gene, thus allowing the measurement of HIV-1 LTR transcription (the protocol recommended by the manufacturer was used). The HIV-1 Tat-expressing plasmid was used to facilitate the activation of HIV-1 transcription. HeLa S3 cells were purchased from ATCC, and were maintained in F-12K media containing 10% fetal bovine serum and penicillin-streptomycin, purchased from Multicell, at 37°C and 5% CO_2. Viable cells were transfected with the LTR-pRL-CMV plasmid, the Tat-expressing plasmid and one of the MCEF constructs using DreamFect (Oz Biosciences) in 80% confluent 6-well plates. The cells were lysed using passive lysis buffer and luciferase assays were performed on each reaction following the protocol recommended by the supplier (Promega, product # E4030), while the activity (luminosity) was quantified using a 20/20 Luminometer (Turner BioSystems).

2.5 RBEIII Mutation Analysis

The hands-on work was done by Juzer Kakal, as follows. The CTG duplication 3' of the RBEIII site was mutated by site-directed mutagenesis on the HIV-1 LTR. The wild-type (monomer) duplication has the sequence **ACTGCTG**ACATC. Two different isogenic HIV-1 mutants, M1 and M4, were generated that had either 1 or 4 base pairs mutated. In M1, wild-type ACTGCTGACATC was mutated to ACTGCTGACAaC; in M4, it was mutated to ACTGCTGAgtag (mutations were based on the study by Sadowski *et al.* [90]). All mutations were generated in pHXB2 plasmids. These mutants, along with the wild-type, were transiently transfected into Jurkat E6-1 T-cells. The cells were pre-seeded in 6 well plates at ~$4x10^6$ cells/1.6mL. The plasmids were prepared for transfection by mixing in separate tubes M1, M4 and wild-type constructs with Serum Free Medium (SFM) according to the following volumes to normalize concentrations: 3.3 µL of wildtype plasmid in 96.7 µL SFM and 2 µL of M1 and M4 plasmids in 98 µL SFM. For a negative control, a blank SFM solution was used and no DNA was added. 8 µL of SuperFect reagent (Qiagen, Cat. no. 301305) was added to each tube and incubated for 10 minutes at room temperature, and the plasmids were subsequently added dropwise to the cells. The samples were incubated overnight at room temperature to allow for recovery and the formation of transfection complexes. To each sample, 2 mL of fresh Jurkat cells ($2x10^6$ cells/mL) were added to each culture and mixed by pipetting. Each culture was split into two 2 mL samples to begin anti CD-3 and phytohemagglutinin (PHA) treatments. The anti-CD3 treatment was carried out using anti-CD3 antibody MACSIBead Particles (Miltenyi Biotec, order no.130-091-441). A stock solution of anti-CD3 beads was made by loading 100 µL anti-CD3 beads onto 500 µL MACSIBead particles. From this stock, 100 µL was washed in SFM and then resuspended in 100 µL SFM. In the wells, each cultured sample contained 900 µL SFM-suspended cells and 100 µL anti-CD3 beads. The final concentration of the beads used was 2.5

μg/mL. The cells were stimulated for 24 hours at room temperature. A stock PHA solution was resuspended in PBS + 0.1% BSA at a concentration of 5 μg/μL. This solution was diluted in SFM to 10 μg/mL and used for the treatment of the cultures. Following treatment, 1 mL samples (from treated and untreated cultures) were removed at the 48 hour time point (1 mL fresh log phase Jurkat was added to the ongoing culture) and 96 hour time point and spun down, where the supernatant was analyzed for p24 using the HIV-1 p24 Capture Elisa Kit from ImmunoDiagnostic, Inc. (product no. 103). The procedure was carried out in accordance with the protocol provided with the kit; no alterations were made. (The manufacturer recommended diluting the supernatant between 1:2 and 1:10 in diluent buffer; in this experiment, a 1:2 dilution was made).

3. RESULTS

3.1 Phylogenetic Analysis

Many sequences for the AF4 family proteins were found in the public GenBank database; some were only partial sequences, and therefore were not used in this analysis. All of the sequences used here were the full sequences of the proteins, not nucleic acids. Human, chimpanzee, monkey, rat, mouse and horse species were analyzed. Although amino acid sequences for chicken, zebra fish, cow and frog species were found for some AFF proteins, only species having all 4 representatives in the data base (AFF1 to AFF4) were included in this analysis. An exception to this was the outgroup from *Drosophila*, a species which appears to have only a single distantly-related AFF member, called Lilliputian [87]. Neighbour joining is a method for constructing phylogenetic trees using DNA or protein sequences. It generates the best tree by using the distances between each protein to the other proteins included in the study; the pair with the shortest difference in distance is linked, or neighbour joined, by a single node (ancestor). The rest of the tree is constructed by removing the newly-linked pair, and using the node instead, as a new protein to which the remaining proteins are compared. The bootstrapping technique provides a means for determining the confidence values for the clades in a phylogenetic analysis. When a bootstrap is performed, the program will randomly select a specific number of sites on the sequences and generate trees using these sites. This process is repeated many times (in this case, 32,000) using different sites and trees, and the number of times the clade was generated is expressed as a percentage of the number of trees produced – the bootstrap value. The results (fig. 3) demonstrate that there are four main clades, one corresponding to each of the AF4 proteins. The first divergence event separated the lineage that gave rise to the AFF1 clade (bootstrap value 100%) from the lineage that gave rise to the AFF4, AFF2 and AFF3 clades. The second lineage was first separated to give rise to AFF4 (bootstrap

value 100%), and another time to give rise to AFF2 and AFF3 (both bootstrap values 100%). The distances of the nodes at divergence points from each of the branches are equal across the board.

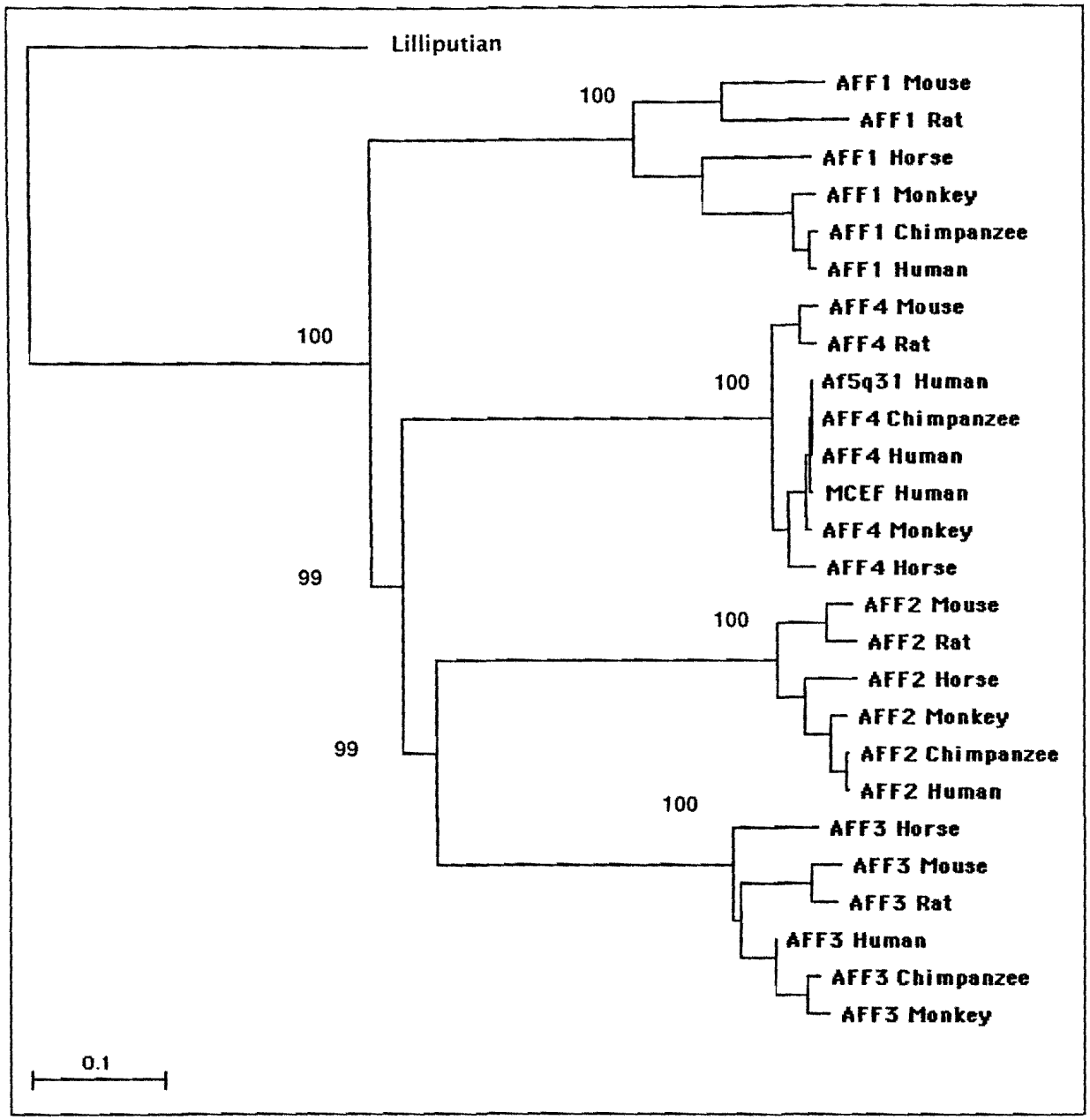


Figure 3. Phylogenetic Tree of AF4 family members. The phylogenetic analysis was performed in MacVector, where ClustalW and Neighbour Joining was used to generate the tree, and the bootstrap values from 32,000 resamplings are provided at each node as percentages. The outgroup used to derive the tree was Fly Lilliputian from *Drosophila*.

The tree also shows that the two proteins AFF2 and AFF3 form sister clades, meaning that these two proteins originate from the same lineage and share an exclusive common ancestor; this is

demonstrated in the tree by a single lineage splitting to form two new ones. The bootstrap value at this node is 99% (fig. 3).

3.2 SDS-PAEG and Western Blots

For the result shown in this paper (fig. 4), the nuclear extract (NE13) that was used had been prepared by Dr. Estable from HeLa S3 cells following the procedure described in Estable *et al.*, 2007 [38, 91]. The samples were run on an SDS-PAGE and then analyzed by Western blotting. Several antisera generated against different regions of the full length AFF4/AF5q31/MCEF protein were used. It is clear from the results that there are a few bands that appear with all of the antisera used, which also happen to be the major bands for the 6 antisera. A band at 17 kDa is very faintly detected by all antisera (except N37). Another band of approximately 40 kDa is detected with all 6 antisera, although it appears to be most efficiently detected by 827, 823 and AFF4. Another band that is seen across the board is at approximately 55 kDa; darker more prominent bands are seen with 823, 826 and N37. The antisera also mutually detected a band at 48 kDa, but is appears the darkest with 823. The 827 antibody detected 3 other major bands, at 140 kDa, 100 kDa and 30 kDa. The band at 140 kDa was also detected by 826, but to a significantly lower extent. This distinct 140 kDa band also appears with 823, AFF4 and N37, but it is absent with N36. Many bands of lesser intensity are detected with N37 that are not detected with the other antisera, specifically a band larger than 225 kDa. The bands detected with 823 are the darkest compared with those detected with other antisera.

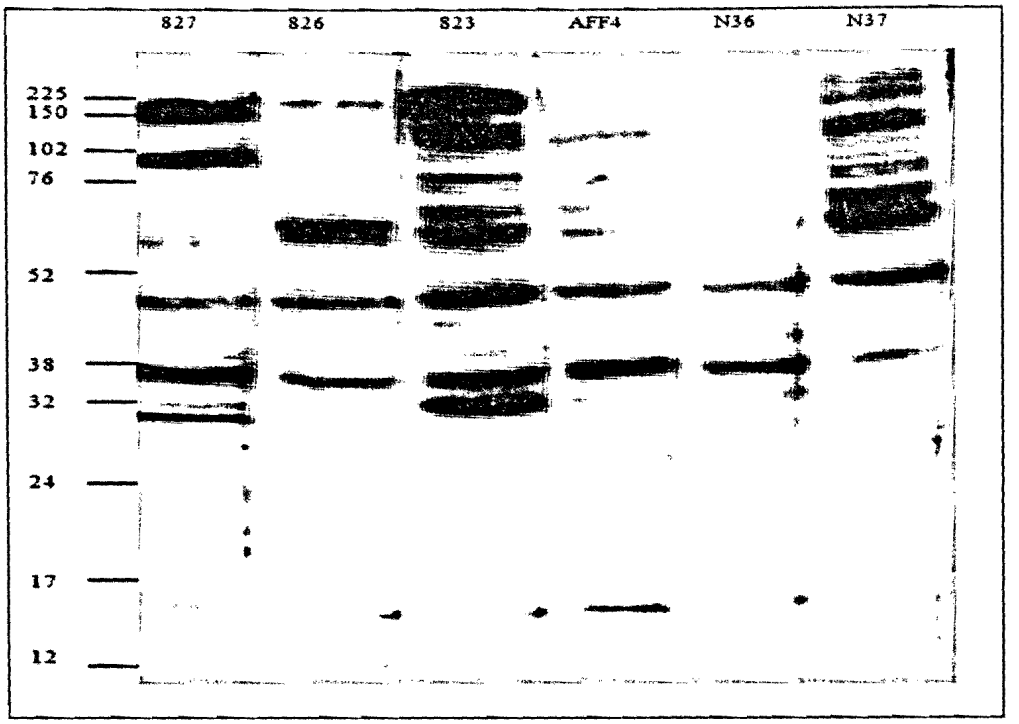


Figure 4. Preliminary results of MCEF western blot analysis. Different MCEF primary antibodies were used during western blotting for each lane. From left to right (in brackets are indicated the amino acids numbers of full length MCEF against which the antisera were generated) – lane 1: 827 (1- 420); lane 2: 826 (1- 530); lane 3: 823 (1- 715); lane 4: AFF4 (189- 206); lane 5: N36 (857- 870); lane 6: N37 (1139- 1154). 20µg of HeLa S3 nuclear extract (NE13) was loaded into each lane.

Western blots were also carried out using other cellular extracts, including HeLa whole cell extracts (+/- alkaline phosphatase). NE13, SCNE, SCWCE and SCWCE+AP samples were run on an SDS-PAGE and a western blot was subsequently performed using 827 primary antibody. The result obtained is shown in fig. 5. At first glance, it is clear that the bands are detected the most strongly in NE13. The banding pattern of NE13 is similar to that seen in fig. 4 (lane 1: 827); the major bands detected were at 140 kDa, 100 kDa, 40 kDa and 30 kDa. Moreover, the banding patterns seen for SCNE and SCWCE are also similar to that of NE13, where the major bands appear at 140 kDa, 100 kDa and 40 kDa but the 30 kDa is absent; the most striking difference is the intensity of the bands, where those detected in NE13 appear more prominent than those detected in SCNE and SCWCE. The 140 kDa and 100 kDa bands were

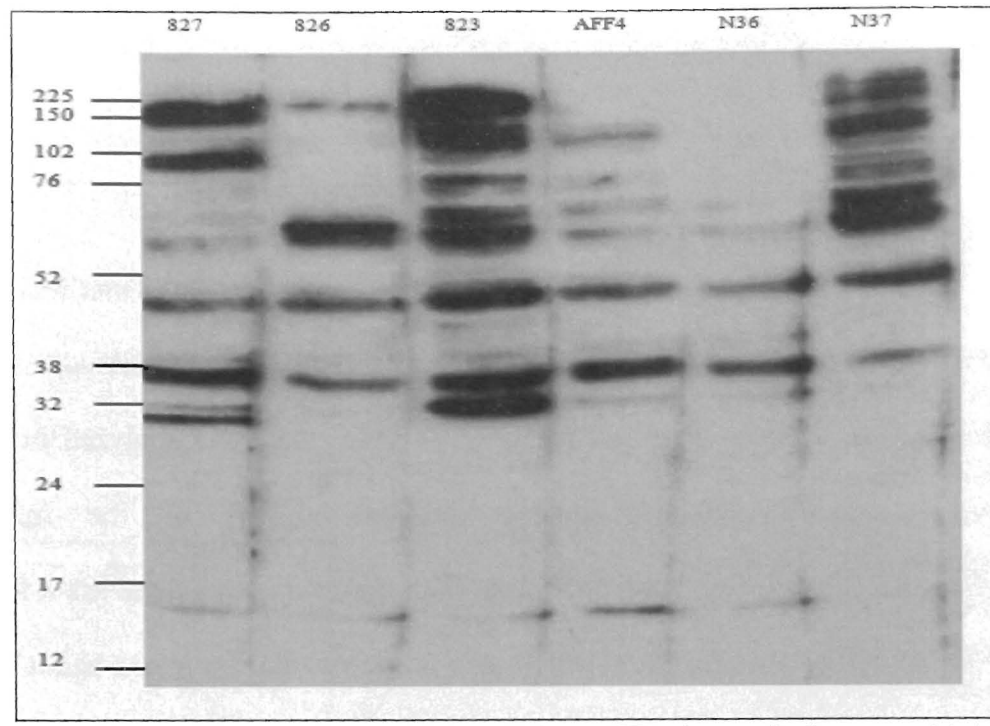


Figure 4. Preliminary results of MCEF western blot analysis. Different MCEF primary antibodies were used during western blotting for each lane. From left to right (in brackets are indicated the amino acids numbers of full length MCEF against which the antisera were generated) – lane 1: 827 (1- 420); lane 2: 826 (1- 530); lane 3: 823 (1- 715); lane 4: AFF4 (189- 206); lane 5: N36 (857- 870); lane 6: N37 (1139- 1154). 20µg of HeLa S3 nuclear extract (NE13) was loaded into each lane.

Western blots were also carried out using other cellular extracts, including HeLa whole cell extracts (+/- alkaline phosphatase). NE13, SCNE, SCWCE and SCWCE+AP samples were run on an SDS-PAGE and a western blot was subsequently performed using 827 primary antibody. The result obtained is shown in fig. 5. At first glance, it is clear that the bands are detected the most strongly in NE13. The banding pattern of NE13 is similar to that seen in fig. 4 (lane 1: 827); the major bands detected were at 140 kDa, 100 kDa, 40 kDa and 30 kDa. Moreover, the banding patterns seen for SCNE and SCWCE are also similar to that of NE13, where the major bands appear at 140 kDa, 100 kDa and 40 kDa but the 30 kDa is absent; the most striking difference is the intensity of the bands, where those detected in NE13 appear more prominent than those detected in SCNE and SCWCE. The 140 kDa and 100 kDa bands were

detected (but less intensely) in SCWCE+AP as well. However, the lower band at 40 kDa was completely absent, but 3 smaller bands, which were not detected in the other 3 samples, appeared consecutively at molecular weights less than 24 kDa (fig. 5). The banding patterns in fig. 4 and fig. 5 were reproduced on other western blots (results not shown).

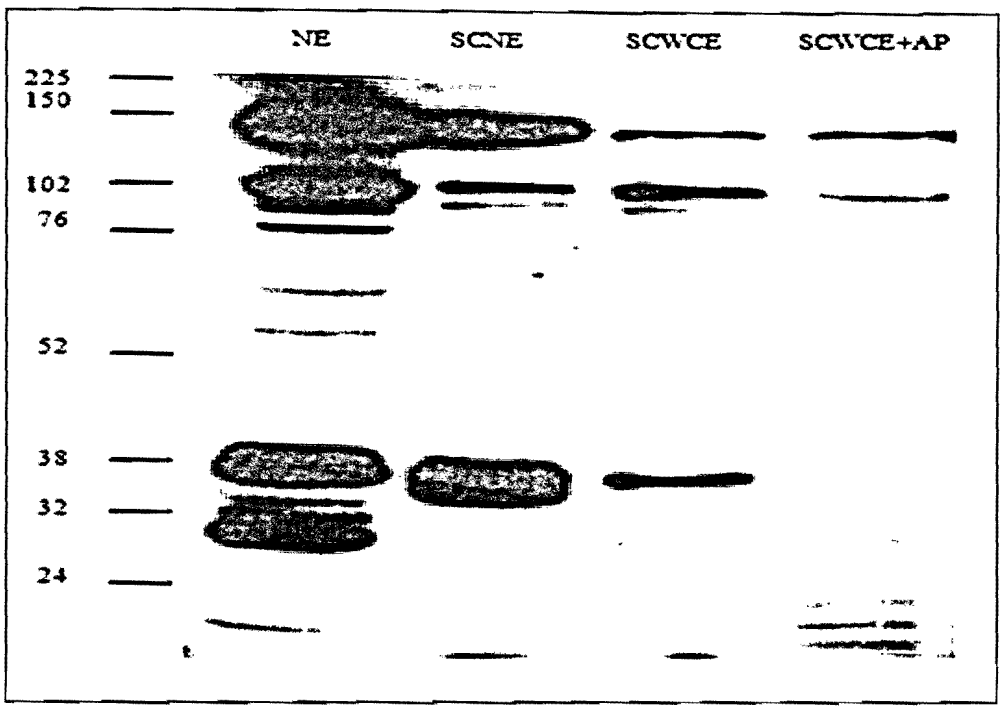


Figure 5. Results of western blot analysis using 827 primary antisera on different HeLa cellular extracts. From left to right – NE: Nuclear Extract 13; SCNE: Santa Cruz Nuclear Extract; SCWCE: Santa Cruz Whole Cell Extract; SCWCE+AP: +Alkaline Phosphatase. 20 µg of each sample was loaded into its' corresponding lane.

3.3 Phosphorylation Experiments

3.3.1 Immunomagnetic Separation

The structure of MCEF was investigated further by first immunomagnetically separating MCEF from HeLa S3 nuclear extracts and then running the nuclear extract, flow-through and elutions on an SDS-PAGE, which was then analyzed by western blotting (anti-MCEF 827 was used as a primary antibody for all western blots in this section). The immunomagnetic separation was actually performed twice; once without any alkaline phosphatase treatment, and another

detected (but less intensely) in SCWCE+AP as well. However, the lower band at 40 kDa was completely absent, but 3 smaller bands, which were not detected in the other 3 samples, appeared consecutively at molecular weights less than 24 kDa (fig. 5). The banding patterns in fig. 4 and fig. 5 were reproduced on other western blots (results not shown).

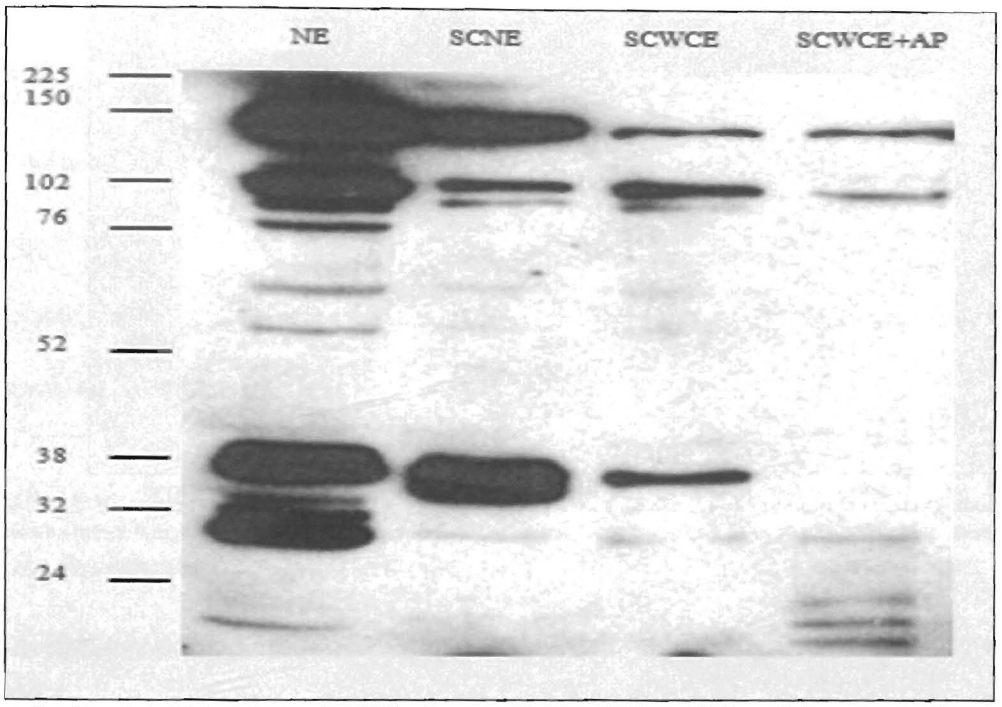


Figure 5. Results of western blot analysis using 827 primary antisera on different HeLa cellular extracts. From left to right – NE: Nuclear Extract 13; SCNE: Santa Cruz Nuclear Extract; SCWCE: Santa Cruz Whole Cell Extract; SCWCE+AP: +Alkaline Phosphatase. 20 µg of each sample was loaded into its' corresponding lane.

3.3 Phosphorylation Experiments

3.3.1 Immunomagnetic Separation

The structure of MCEF was investigated further by first immunomagnetically separating MCEF from HeLa S3 nuclear extracts and then running the nuclear extract, flow-through and elutions on an SDS-PAGE, which was then analyzed by western blotting (anti-MCEF 827 was used as a primary antibody for all western blots in this section). The immunomagnetic separation was actually performed twice; once without any alkaline phosphatase treatment, and another

time with alkaline phosphatase treatment of the nuclear extract NE13.

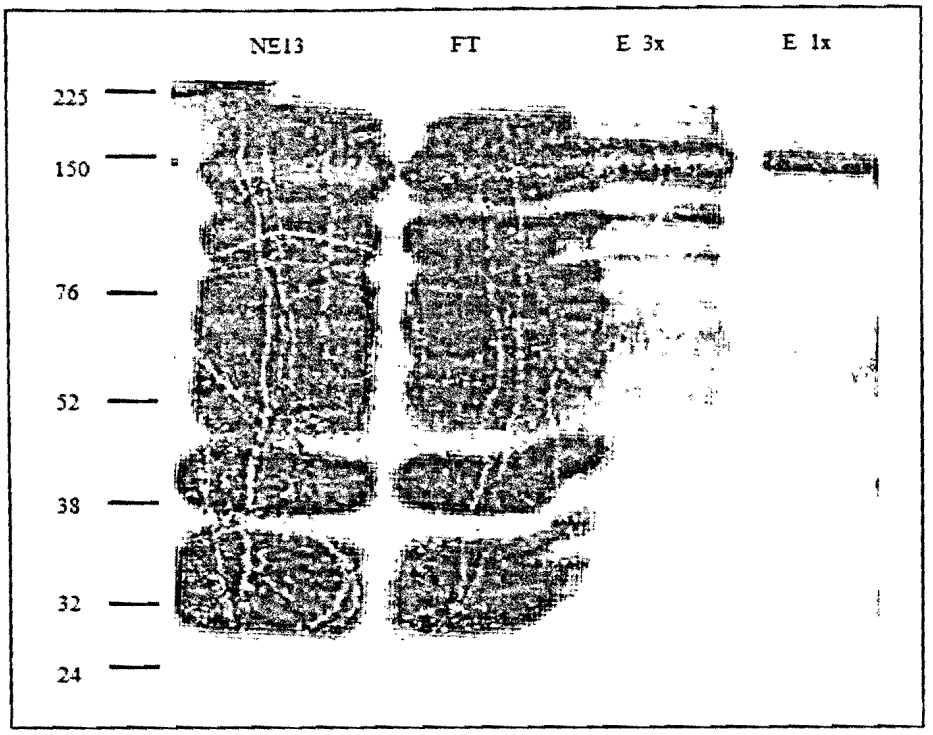


Figure 6. Immunomagnetic separation of HeLa S3 Nuclear Extract NE13. NE13: HeLa S3Nuclear extract NE13; FT: flow through; E: first elution with citrate buffer; 3X: three times equivalent amount loaded; 1X: equivalent amount to original sample loaded.

The western blot in fig. 6 shows the results of the immunomagnetic separation that was performed without alkaline phosphatase treatment. The results of probing NE13 with 827 anti-MCEF antibody obtained in the first two experiments were confirmed again in this experiment; the same major bands are detected at 140 kDa, 100 kDa, 40 kDa and 30 kDa. The same banding pattern is seen in the flow through lane as that in the NE13 lane. However, the two elution lanes, E 3x and E 1x, exhibit different banding patterns. The only major band that is detected in E 3x is at 140 kDa, but was detected weaker than it was in the NE13 and flow-through samples. A smaller, significantly weaker band is also detected at 100 kDa. The 140 kDa is also detected in E 1x at a slightly weaker intensity than E 3x, but the 100 kDa is very weakly detected (weaker than E 3x).

time with alkaline phosphatase treatment of the nuclear extract NE13.

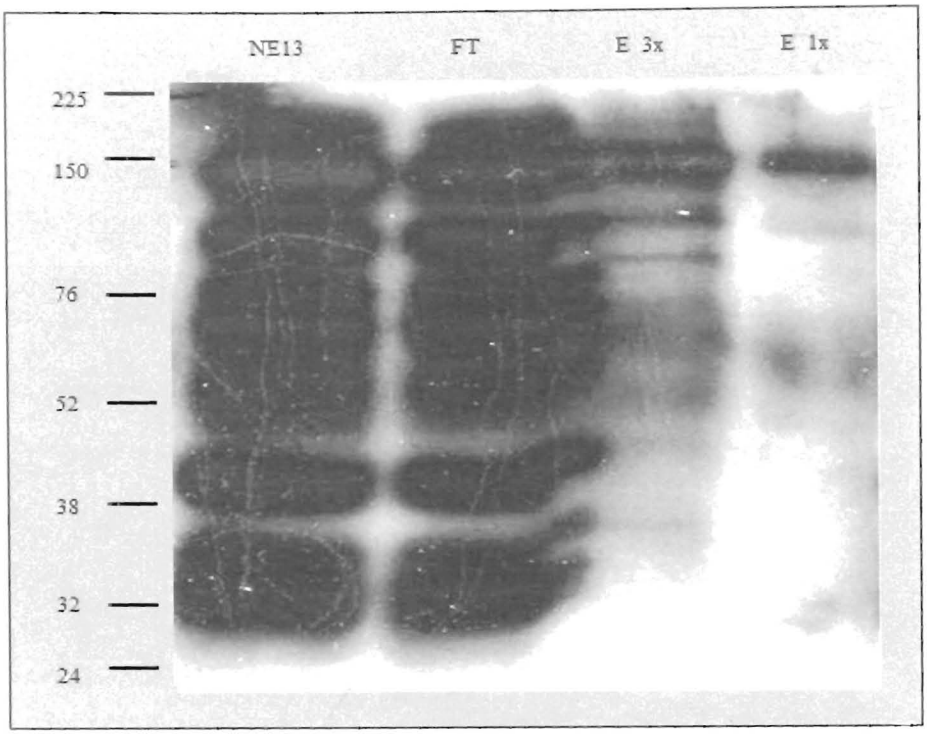


Figure 6. Immunomagnetic separation of HeLa S3 Nuclear Extract NE13. NE13: HeLa S3Nuclear extract NE13; FT: flow through; E: first elution with citrate buffer; 3X: three times equivalent amount loaded; 1X: equivalent amount to original sample loaded.

The western blot in fig. 6 shows the results of the immunomagnetic separation that was performed without alkaline phosphatase treatment. The results of probing NE13 with 827 anti-MCEF antibody obtained in the first two experiments were confirmed again in this experiment; the same major bands are detected at 140 kDa, 100 kDa, 40 kDa and 30 kDa. The same banding pattern is seen in the flow through lane as that in the NE13 lane. However, the two elution lanes, E 3x and E 1x, exhibit different banding patterns. The only major band that is detected in E 3x is at 140 kDa, but was detected weaker than it was in the NE13 and flow-through samples. A smaller, significantly weaker band is also detected at 100 kDa. The 140 kDa is also detected in E 1x at a slightly weaker intensity than E 3x, but the 100 kDa is very weakly detected (weaker than E 3x).

3.3.2 Alkaline Phosphatase Assay

The alkaline phosphatase treatment was carried out using Calf Intestinal Alkaline Phosphatase (CIAP) that is commercially available. CIAP works by removing the phosphate group of a protein, thereby allowing for the identification of phosphorylated isoforms of the protein of interest, MCEF. In a preliminary experiment, 2 μL NE13 was assayed for consecutively increasing amounts of CIAP (several CIAP enzymes were used that were purchased from Invitrogen, Fermentas and New England Biolab) and loaded onto an SDS-PAGE, followed by a western blot analysis (results not shown). The results showed that the major band detected was 140 kDa, where the intensity of the band was the same for different CIAP concentrations and different CIAP enzymes. The results also indicated that a 100 kDa was also detected, but to a lesser extent than the 140 kDa.

The final experiment performed combined immunomagnetic separation and alkaline phosphatase treatments into a single experiment. The NE13 nuclear extracts were first treated with a CIAP following the procedure outlined in section 2.3.2, and subsequently immunomagnetically separated, run on an SDS-PAGE and examined further by western blotting. The results for this experiment are shown in fig. 7. The NE13 and the flow-through lanes demonstrated the same banding pattern, with the major bands detected at 140 kDa, 100 kDa and 40 kDa. However, for the most part, the signal detected for the bands in the flow through lane appeared less strong than the same bands detected in the NE13 lane. No bands were detected in the first wash, second wash and third wash lanes. The three major bands were detected in the first elution at the same molecular weight and intensity as those in the first two lanes, at 140 kDa, 100 kDa and 40 kDa. In the second elution lane, only the 140 kDa band appears to be a major band; a band at 100 kDa is also detected but only weakly.

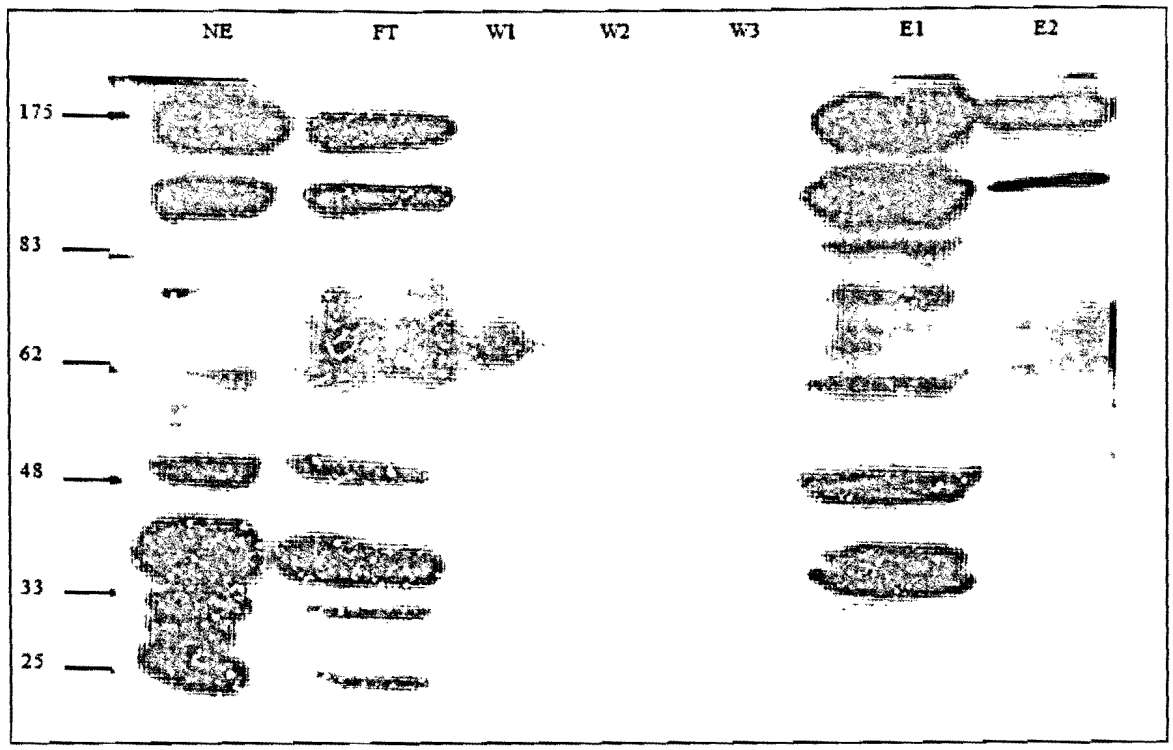


Figure 7. Results of western blot analysis of immunomagnetic separation of MCEF following treatment with alkaline phosphatase. From left to right - NE: Nuclear Extract 13; FT: flow though; W1: first wash; W2: second wash; W3: third wash; E1: first elution; E2: second elution. (5 times the amount of sample loaded into the first five lanes was loaded into the E1 and E2 lanes).

3.4 MCEF Deletion Analysis

Four different MCEF plasmids were prepared, each expressing different portions of the MCEF ORF, driven by the CMV promoter sequences. The CMV promoter is recognized by the cellular RNA polymerase II transcription machinery and drives the strong expression of any ORF cloned downstream of it [38, 51]. The CMV MCEF-expressing constructs were co-transfected into HeLa cells, along with a CMV-HIV-1 Tat-expressing construct and a reporter construct. The reporter construct consisted of the HIV-1 LTR promoter region, cloned in-front of the *Renilla* luciferase gene. Thus, the transcription level from the HIV-1 LTR, in the presence of HIV-1 Tat, could be measured by simply lysing cells and measuring luminosity [51].

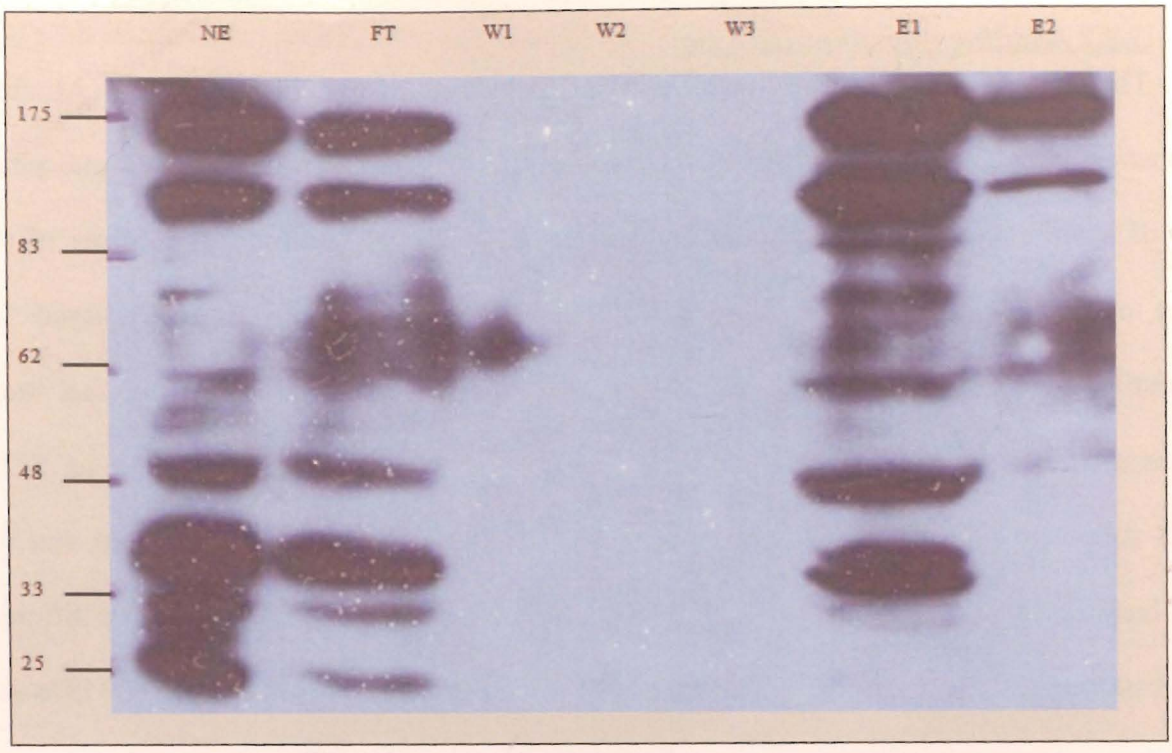


Figure 7. Results of western blot analysis of immunomagnetic separation of MCEF following treatment with alkaline phosphatase. From left to right - NE: Nuclear Extract 13; FT: flow though; W1: first wash; W2: second wash; W3: third wash; E1: first elution; E2: second elution. (5 times the amount of sample loaded into the first five lanes was loaded into the E1 and E2 lanes).

3.4 MCEF Deletion Analysis

Four different MCEF plasmids were prepared, each expressing different portions of the MCEF ORF, driven by the CMV promoter sequences. The CMV promoter is recognized by the cellular RNA polymerase II transcription machinery and drives the strong expression of any ORF cloned downstream of it [38, 51]. The CMV MCEF-expressing constructs were co-transfected into HeLa cells, along with a CMV-HIV-1 Tat-expressing construct and a reporter construct. The reporter construct consisted of the HIV-1 LTR promoter region, cloned in-front of the *Renilla* luciferase gene. Thus, the transcription level from the HIV-1 LTR, in the presence of HIV-1 Tat, could be measured by simply lysing cells and measuring luminosity [51].

The results of this deletion analysis experiment are shown in figure 8. The highest level of luciferase expression was recorded by the cells transfected with the plasmid containing the middle portion of the MCEF protein, amino acids 402- 745 (blue bar). This level was even slightly higher than the luciferase level reported by the cells transfected with the plasmid containing the full sense (reverse) sequence of MCEF (purple bar). The lowest level of luciferase protein was found in the cells expressing only the C-terminal of MCEF, amino acids 694- 1163 (green bar). The expression of luciferase was slightly higher in the cells expressing only the N-terminal region (amino acids 1- 715) of MCEF than that seen with C-terminal MCEF. The raw data generated are shown in table 2-A.

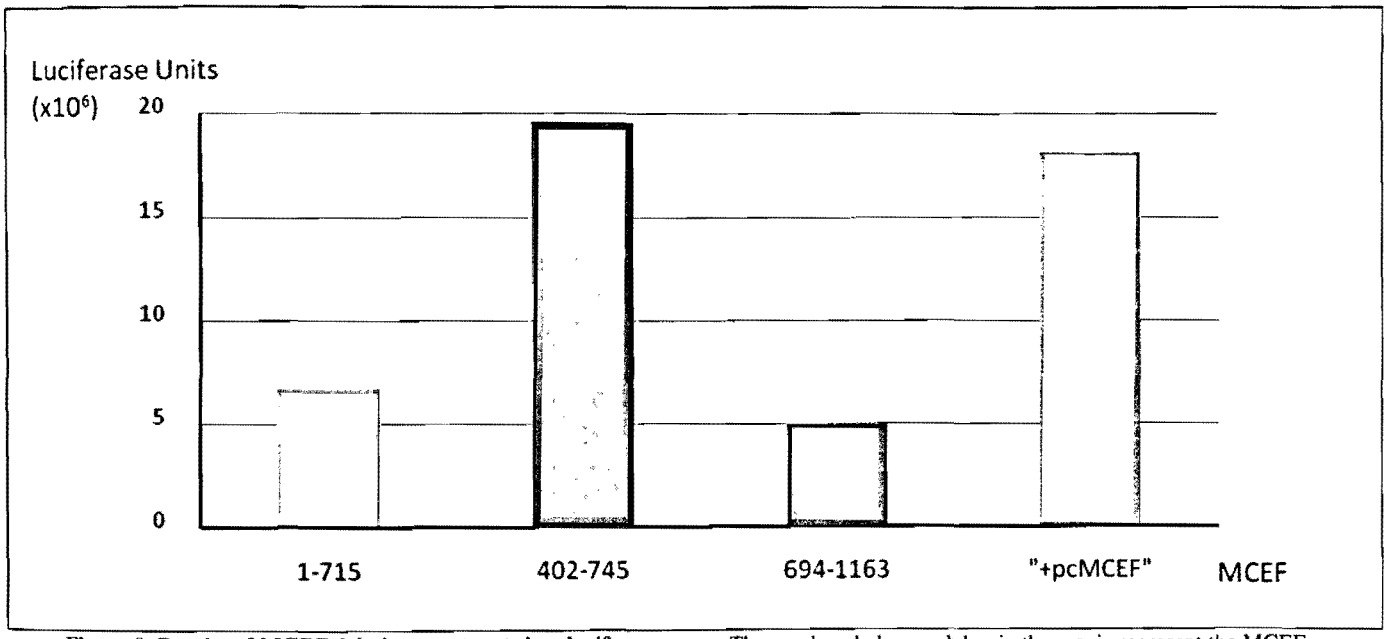


Figure 8. Results of MCEF deletion constructs in a luciferase assay. The numbers below each bar in the x-axis represent the MCEF amino acid numbers (of the full-length protein) included in the corresponding construct. Orange bar: amino acids 1- 715 only; blue bar: amino acids 402- 745 only; green bar: amino acids 649- 1163; purple bar: sense MCEF (reverse full MCEF sequence).

The results of this deletion analysis experiment are shown in figure 8. The highest level of luciferase expression was recorded by the cells transfected with the plasmid containing the middle portion of the MCEF protein, amino acids 402- 745 (blue bar). This level was even slightly higher than the luciferase level reported by the cells transfected with the plasmid containing the full sense (reverse) sequence of MCEF (purple bar). The lowest level of luciferase protein was found in the cells expressing only the C-terminal of MCEF, amino acids 694- 1163 (green bar). The expression of luciferase was slightly higher in the cells expressing only the N-terminal region (amino acids 1- 715) of MCEF than that seen with C-terminal MCEF. The raw data generated are shown in table 2-A.

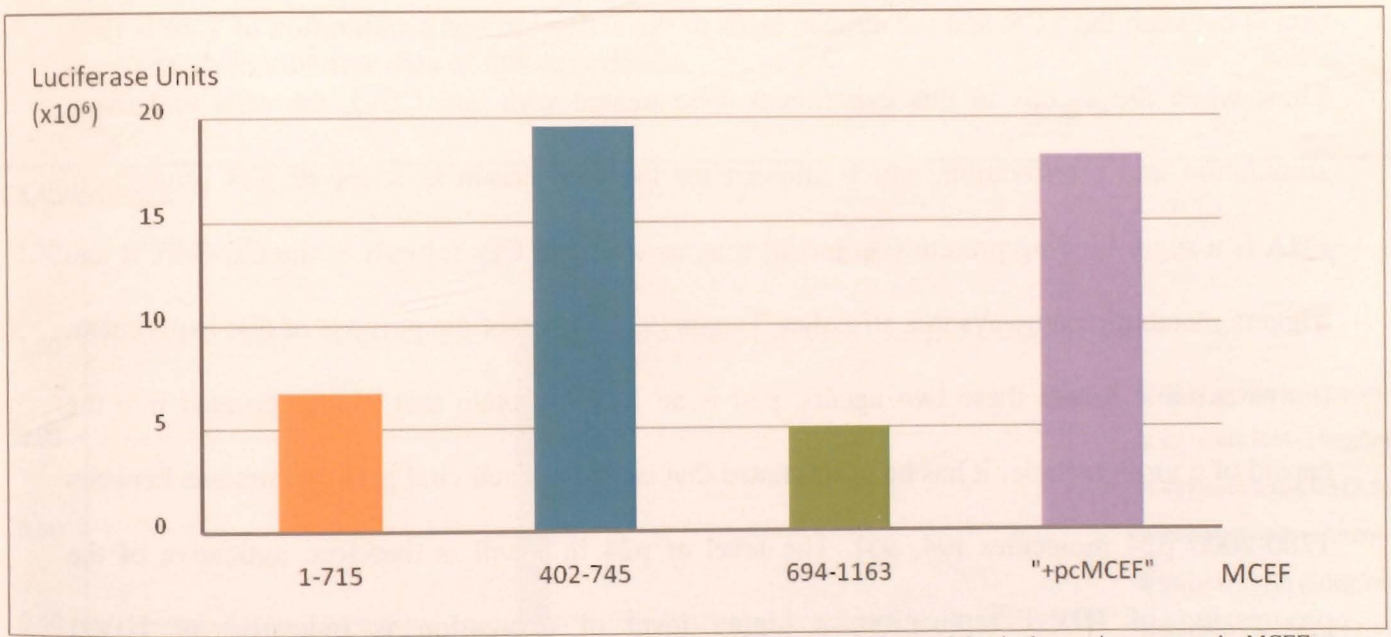


Figure 8. Results of MCEF deletion constructs in a luciferase assay. The numbers below each bar in the x-axis represent the MCEF amino acid numbers (of the full-length protein) included in the corresponding construct. Orange bar: amino acids 1- 715 only; blue bar: amino acids 402- 745 only; green bar: amino acids 649- 1163; purple bar: sense MCEF (reverse full MCEF sequence).

3.5 RBEIII Mutation Analysis

In this experiment, the RBEIII site of the HIV-1 LTR was mutated by site directed mutagenesis. Two isogeneic mutants were generated, M1 and M4. These constructs were then transiently transfected into Jurkat cells, which were either treated or untreated with two different T-cell stimulants, anti-CD3 antibodies and PHA, the level of p24 expression was then measured in each reaction. Jurkat cells are a specific type of T-cells that are characterized by exhibiting immortal properties; this cell line is particularly suitable for the study of HIV-1 because the progression of the virus can be measured. The rationale behind using anti-CD3 antibodies is that CD3 is a component of T-cell receptors (TCR). T-cells respond to an external signal when the TCR recognize an antigen presented by an antigen presenting cell (APC). The interaction that occurs between the TCR and the antigen leads to the activation and proliferation of T-cells [92]. Thus, when the T-cells in this experiment were treated with anti CD-3, the cells underwent stimulation and proliferation, which allowed for the comparison in terms of p24 production. PHA is a sugar-binding protein (i.e. lectin) that, as with anti CD-3, binds to the CD-3-TCR and triggers metabolic pathways that stimulate T-cells [93]. Thus, for the purpose of this experiment, it was suitable to use these two agents. p24 is an HIV-1 protein that is incorporated into the capsid of a virus particle. It has been estimated that capsid of each viral particle contains between 1200-2000 p24 molecules [94, 95]. The level of p24 in a cell is therefore indicative of the progression of HIV-1 replication; a higher level of expression is indicative of HIV-1 proliferation. This makes p24 an excellent target for the measurement of HIV-1 development.

The mock constructs did not contain any DNA. With or without treatment, no p24 protein was detected when these constructs were transfected into Jurkat cells. These results were consistent at the 48-hour and 96-hour time points. The expression of p24 in the untreated

monomer (wild-type) construct doubled between the 48-hour and the 96-hour time points. p24 expression also doubled in the PHA-treated reaction, compared to the untreated reaction, after 48 hours. Moreover, there was a 3.5-fold increase in the absorbance in the PHA-treated cells between the 48-hour and 96-hour time points. The levels of p24 detected in the anti-CD3-treated cells were approximately the same as the levels in the PHA-treated cells, 48 hours following treatment. The expression of p24 was constant between 48 hours and 96 hours in the anti-CD3 treated cells. There were no significant differences in p24 expression in the 48-hour PHA-treated, 48-hour-anti-CD3, 96-hour untreated and 96-hour anti-CD3 treated cells. The results generated from the M1 and M4 constructs were almost identical; virtually no p24 was detected in either construct, with or without T-cell stimulation at both time points (fig. 9). Table A-3 in the appendix shows the raw data of this experiment.

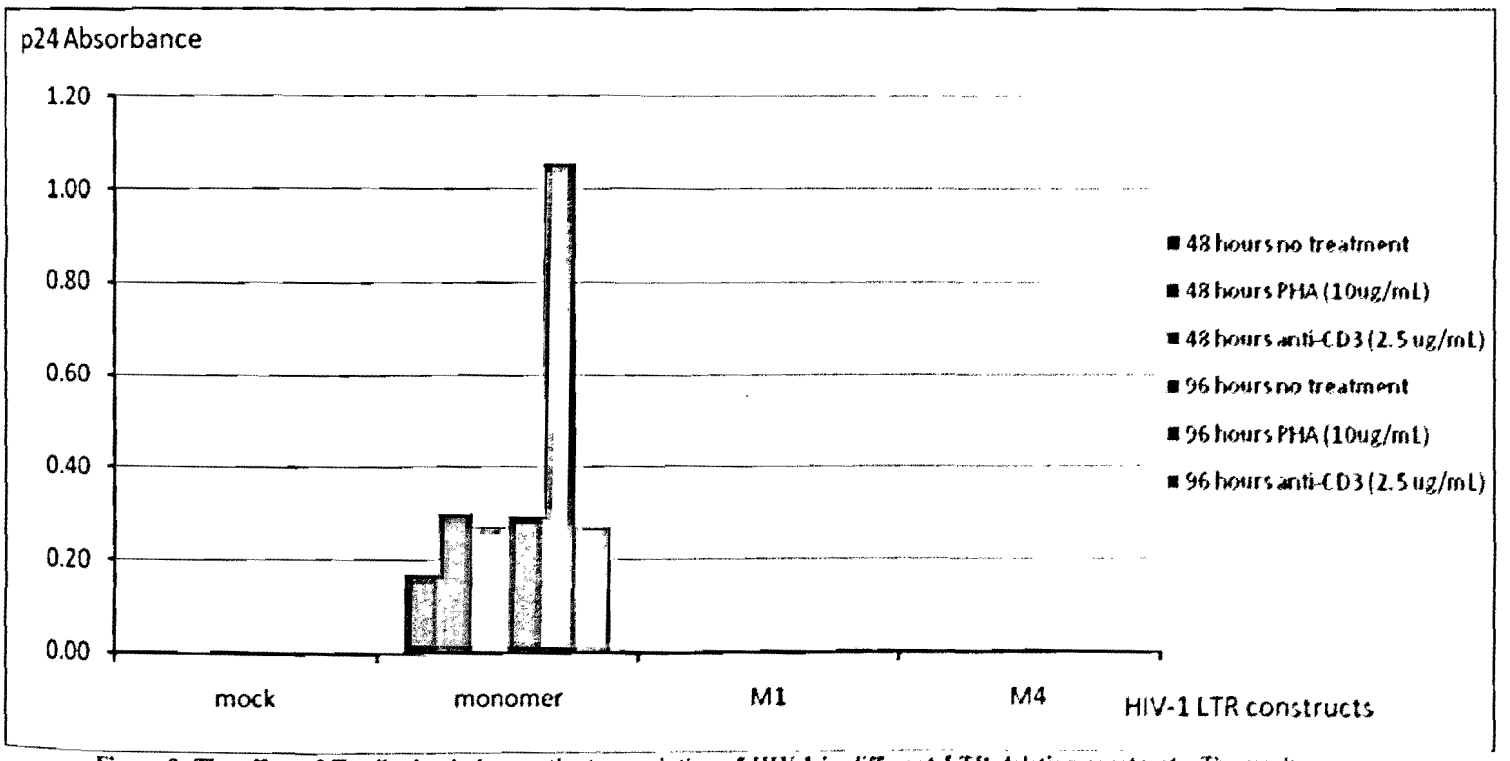


Figure 9. The effect of T-cell stimulation on the transcription of HIV-1 in different LTR deletion constructs. The results were based on three replicas. The mock construct contains no DNA; monomer contains the wildtype CTG construct (ACTGCTGACATC); mutant M1 contains 1 bp mutation (ACTGCTGACAaC); mutant M4 contains 4 bp mutations (ACTGCTGAgtag).

4. DISCUSSION

4.1 Phylogenetic Analysis

The purpose of conducting this analysis was to examine the evolutionary relatedness of the AF4 proteins from their common ancestor and to determine the relative similarity among the four proteins. The results suggest a first divergence gave rise to AFF1 and to the lineage that eventually gave rise to AFF4, and later AFF2 and AFF3. The outgroup used was Lilliputian (Lilli), and the rationale behind that was the similarity it shares with the AF4 proteins, as demonstrated by Wittwer *et al.* [87]. Lilli is the only protein in *Drosophila* that seems to be related to the AF4 proteins, as confirmed by sequence similarities; it shares many of the homology domains observed in the AF4 proteins (fig. 2). The function of Lilli in *Drosophila* has not yet been deciphered, but studies have indicated that it is necessary for normal growth and cell identity [87]. Lilli was the only sequence used that belonged to the phylum *Arthropoda*; all the other sequences belonged to the superphylum *Deuterostomia.* Using the results obtained from the tree, it would be safe to assume that the ancestor of AFF *Deuterostomia* developed at approximately the same time as *Bilatera*, the common ancestor that gave rise to *Arthropoda and Deuterostomia* [51].

Moreover, the results discussed in section 3.1 suggest that AFF2 and AFF3 share a more recent common ancestor than either does with AFF1 and AFF4; meaning that AFF2 and AFF3 are more closely related to each other than to the remaining 2 proteins. Similarly, this would suggest that AFF2, AFF3 and AFF4 also share a more common ancestor with each other than with AFF1, implying that the former 3 proteins are more closely related to one another than they are to AFF1. It is also important to note that the distances of the nodes from the branches demonstrate whether the proteins diverging at this point are differentially related to the protein

they diverged from. In this case, all the distances were equal, which suggests that the AFF4 may be equally related to AFF2 and AFF3. Likewise, AFF2, AFF3 and AFF4 are all equally related to AFF1. Taking this in conjunction with the fact that 3 of the 4 AFF proteins are MLL fusion proteins, it is likely that, as *Bilatera* evolved, the four AFF proteins developed from a single sequence by ancestral duplication [51, 96]. Although this analysis was carried out previously by Niedzielski *et al.* [51], the results did not completely overlap the results generated here. They showed that not only do AFF2 and AFF3 form sister clades, but AFF1 and AFF4 do so as well. The discrepancy between the two sets of results may have been a consequence of using an extra set of sequences that belong to the horse species. Increasing the number of species in an analysis will provide more data and insight on the proteins in question and therefore enhance the reliability of the study. Additionally, some of the sequences used in the previous study were partial sequences. Many of these sequences were found either completely erased or updated to the full versions upon searching for them in this study (the following sequences were used to perform in the initial analysis in 2007 and were updated, replaced or removed from the database: XP_223161.4, NP_032058.1, XP_001054673.1, XP_525831.1, AAH52061.1, XP_343560.2, XP_517928.1, and XP_220420.2).

4.2 SDS-PAGE and Western Blot Analysis

Performing these western blots was an important step for characterizing the anti-MCEF antisera available in the lab and for learning some preliminary information about the structure of MCEF that can be further analyzed and broken down into separate areas of research. In the first experiment (fig.4), nuclear extract NE13 was used. As can be seen from the results, some antisera detected bands that other antisera did not. More importantly, all antisera were able to detect peptides of molecular weights 17 (except N37), 40, 48 and 55 kDa. Because these are only

partial peptides of the full-length MCEF protein, they can either be degradation products of MCEF or as splice variants of MCEF generated by alternative splicing of MCEF mRNA. The antisera used are generated against several exons of MCEF: 827, 826 and 823 together span exons 1- 11; AFF4 was generated against a 17 amino acid stretch in exon 3; N36 recognizes amino acids 857- 870 in exon 13; N37 was generated against amino acids 1139- 1154 in exon 21 (a sequence conserved among all 4 AF4 proteins, therefore not unique to MCEF). Given that the 17, 40, 48 and 55 kDa bands are detected with all antisera (except N37), this means that these peptides must contain partial sequences of each of the exons against which the antisera was generated, namely exons 3, 13 and 21. Also, because the molecular weights of the four peptides are different, this must mean that certain sequences that are included in the 55 kDa are absent in the smaller (17 kDa, 40 kDa, 48 kDa) peptides. The difference in the intensity of the bands may be attributed to how much of the original antisera sequence is conserved in the sequence of the peptide it is binding to; the lower the intensity of the band, the lesser the amino acids that can be recognised by the antisera. Following the same logic, the 17 kDa was possibly not detected with N37 is because it lacked sequences from exon 21 (which also explains its' relatively small size). Interestingly, the 55 kDa could be the same protein that was detected by Estable *et al.* as part of the study on P-TEFb that led to the discovery of MCEF, where it was thought to be a degradation product or a splice variant of MCEF, similar to the suggestion stated earlier [38].

The results for NE13 obtained with the 827 antisera (both fig.4 and fig.5) yielded four major bands at 140, 100, 40 and 30 kDa. The 140 kDa band corresponds to the full length MCEF protein. As mentioned previously, the theoretical weight of MCEF is 127 kDa, but it is instead consistently detected in experiments at 140 kDa, most likely due to post-translational modifications that impede the migration of the protein through the gel. Full length MCEF is also

detected by 823, AFF4 and N37, demonstrated by the band that appears at approximately 140 kDa in the corresponding lanes. The 826 antibody also detects full length MCEF, but the intensity of the band was weaker than the bands in the remaining lanes. This weaker signal could be attributed to experimental error, as a significantly stronger signal was detected at 140 kDa with 826 in all other western blots performed (results not shown). The 140 kDa was absent when N36 was used. As mentioned previously, the N36 antibody was generated by Biosynthesis Inc. according to the results obtained from a program that identifies the most antigenic regions of MCEF. However, despite using 1:500 dilutions (cf. 1: 1,000), major bands consistently appeared weaker with N36. The antibody itself may not have been a very specific to the region it was generated against, resulting in compromised detection.

The 100 kDa, 40 kDa and 30 kDa bands detected by 827 could be possible isoforms of MCEF that all contain partial regions of the MCEF C-terminal, because they are all detected by an antibody generated against the C-terminal. The latter statement would also eliminate the possibility of the two larger bands (100 kDa and 40 kDa) being products of proteolytic cleavage of full length MCEF (given that their molecular weights add up to 140 kDa).

The N37 antibody was also generated by Biosynthesis by the same approach that was used to generate N36. The region that was identified as most antigenic in exon 21, however, was not specific to MCEF; it was in fact conserved among the AF4 proteins as part of the carboxy-terminal homology domain (fig. 2). N13 is a nuclear extract that contains a multitude of proteins localized to the nucleus, including the AF4 family proteins. Therefore, when N37 was used to detect only MCEF-specific peptides, many bands were detected in the process; the 15 amino acids peptide that the antibody can recognize is not only present in MCEF, but also in AFF1, AFF2 and AFF3, giving rise to the extra bands seen in the N37 lane. Moreover, a band larger

than 225 kDa was stable in all western blots performed with N37. The most probable explanation of this band is non-specific binding or unreactivity, but given that the MLL-MCEF chimeric protein is larger than 225 kDa, it was intriguing to determine whether or not this band corresponds to the chimeric protein. A few experiments (western blots, for the most part) were performed to elucidate an answer; unfortunately, no conclusive result was generated.

The bands detected with the 823 antisera were significantly darker than these detected with the other antisera. The 823 antibody is the longest of the antisera used in this experiment; it was generated against amino acids 1- 715 of MCEF. This difference in length could explain the difference in intensity. These extra amino acids closer to the C-terminal of MCEF may cause the regions in the middle of the antibody to bind more strongly (and less transiently) to the corresponding sequences on the MCEF protein, resulting in a stronger signal.

The second part of the western blot analyses was performed to identify differences in MCEF expression in a variety of cellular extracts, rather than with different antisera. The same primary antibody was used in all lanes, 827. The banding pattern seen in the far-left lane, NE13, in fig. 5 is a replica of the banding pattern seen in the far-left lane, 827, in fig. 4. The same major bands, 140 kDa, 100 kDa, 40 kDa and 30 kDa are seen in these two lanes. Again, the 140 kDa here corresponds to the full length MCEF protein; whereas the latter 3 are hypothesized to be isoforms of MCEF that all partially encompass the C-terminal. The 140 kDa, 100 kDa and 40 kDa bands are also seen in the next lane, SCNE, suggesting that the full length MCEF and its' potential isoforms were all present in the lysate. Although the 30 kDa band is detected with 827 antisera in NE13 nuclear extract, it does not appear when 827 is used for commercially available nuclear extract. NE13 was extracted from cells that were grown in suspension, whereas SCNE was prepared by Santa Cruz Biotechnology from cells that were grown adherently (according to

the protocol outlined by Dignam *et al.* [91]). Usually, cells that are grown in suspension express proteins at higher levels than cells that are grown adherently. This phenomenon could explain the absence of the 30 kDa band in SCNE; it may not have been expressed at all in the cells, or expressed at levels too low to be detected. The level of expression of the full length MCE band at 140 kDa is considerably lower in the whole cell lysate lanes, SCWCE and SCWCE+AP. This is to be expected considering the fact that MCEF is localized to the nucleus; it will be expressed at higher levels in concentrated samples of nuclear extracts than in more diluted samples of whole cell lysates. This trend also applies to the 40 kDa band in the SCWCE lane, where the band appears to be of lesser intensity compared to that seen in SCNE. The 100 kDa band, on the other hand, seems just as faint in SCNE as it does in SCWCE and SCWCE+AP, a result which could also be the consequence of growing the cells in suspension as opposed to adherently, or to alternative splicing in these cells that led to the elimination of a key nuclear localization signal.

However, the disappearance of the 40 kDa band in the SCWCE+AP lane remains the most prominent observation in fig. 5. Three additional bands of consecutively lower molecular weights appeared instead. The addition of alkaline phosphatase (AP) to lysate is usually done to study the pattern of modification of proteins by phosphorylation. AP is responsible for the removal of a phosphate group from a protein residue. The disappearance of the 40 kDa band upon AP treatment suggests that the peptide is phospohrylated. As discussed in section 1.4, the removal of a phosphate group could result in a protein conformational change, which when happens in western blots, leads to the inability of an antibody to recognize the protein (and the lack of a signal). The change in conformation of MCEF due to dephosphorylation could also have marked the protein for degradation, explaining the three smaller bands that appear in its place upon AP treatment.

4.3 Phosphatase Experiments

The results obtained from the last western blot experiment (section 4.2) suggested that the MCEF protein may be phospohrylated. In addition, several papers that verify the phosphorylation of serine and tyrosine residues in MCEF have been published (section 1.4). The purpose of the phosphorylation experiments was to characterize any phospohrylated MCFE isoforms. To achieve this, two different experiments were used: immunomagnetic separation and alkaline phosphatase assay.

4.3.1 Immunomagnetic Separation

Immunomagnetic separation takes advantage of the specificity between a protein and an antibody generated against it to obtain a concentrated, pure sample of the protein that can be used in downstream analysis. In this experiment, an immunomagnetic separation was performed to generate a highly enriched pure MCEF sample that was run on an SDS-PAGE and analyzed by western blotting. Generating an enriched sample of MCEF was also a crucial step in the next part of this analysis, the alkaline phosphatase treatments.

The results (fig. 6) in the NE13 lane were identical to many of the previous experiments discussed. The major bands were detected at 140 kDa, 100 kDa, 40 kDa and 30 kDa, which correspond to full-length MCEF and potential isoforms of MCEF, respectively. The same bands were detected in the flow-through lane. The flow-through theoretically should contain any cellular material that was not bound to the bead-linked antibody (827), which in this case would consist of any cellular material that is not recognized as MCEF, and any unbound MCEF. The four same bands were detected in the flow-through lane, which means that some of the MCEF protein was unbound to the antibody. This could be the result of antibody saturation, where all the antigen-binding sites on the antibody were occupied by MCEF protein, and any excess

MCEF was left in the flow-through. The elution washes theoretically must contain only purified MCEF protein; this was achieved in this experiment. The elution sample E 1x contained full-length MCEF, as reflected by the presence of a 140 kDa band in the corresponding lane. This band, however, was significantly weaker in intensity, suggesting that only a small quantity of full-length MCEF was eluted. This could be due scarcity of antibody; only so many antibody molecules were available for binding to MCEF, which meant that only so much of MCEF was bound and available for elution. This explanation would also account for the relatively larger amount of MCEF that was unbound and washed in the flow-through. Moreover, the incubation performed to bind the cross-linked antibody to the protein may not have been very successful. Although the supplier recommended special incubation apparatus that achieve binding by a series of rotation, mixing and turning cycles, another less expensive method was employed in this experiment; it entailed rotating and turning the sample by attaching the tube to the second hand of a clock. This method for incubation may have resulted in less efficient binding of the protein to the antibody, yielding less MCEF protein in the elution samples.

The 3x elution sample contained three times the amount of elution sample that was loaded into the 1x elution sample. Comparing the intensity of the 140 kDa bands reveals that the amount of MCEF protein that was purified in the 3x elution sample was almost equivalent to that seen in the NE13 and flow-through samples. An extra band at 100 kD3aa was also detected in the 3x elution sample that was not detected in the 1x sample. Although present, this band is significantly weaker than the 100 kDa detected in the NE13 and flow-through samples and the 140 kDa detected in the 3x sample. The lesser detection could be attributed to a naturally lower level of expression of the 100 kDa isoform in the nucleus. The antibody may also have a lower affinity for smaller versions of MCEF that may fold into different 3D confirmations that cause it

to be unrecognized by the antibody; this includes the 40 kDa band that was not detected in the two elution samples. Given that the 40 kDa isoform of MCEF interacts with other proteins (P-TEFb, for example), it may then also be rational to propose that the 40 kDa isoform was not recognized by the antibody because it was in a complex with another protein, resulting in an absence of signal on the western blot.

4.3.2 Alkaline Phosphatase Assay

The previous technique of immunomagnetic separation was used in combination with an alkaline phosphatase treatment of nuclear extracts to ensure that the results generated in figure 5 can be reproduced and can be considered reliable. The pure sample of MCEF produced in immunomagnetic separation gives rise to more accurate results because it eliminates the possibility of non-specific binding and it ensures that any band detected in a western blot following treatment with calf intestinal alkaline phosphatase will be MCEF, in one of its many forms.

The bands detected in the NE13 and flow-through samples in this experiment (fig. 7) were almost identical to the bands seen in the NE13 and flow-through samples in the first immunomagnetic separation (fig. 6). The 140 kDa, 100 kDa and 40 kDa peptides are the major MCEF isoforms detected. The three wash lanes, W1, W2 and W3, contained no MCEF at all because by this stage any MCEF that was not bound to the beads were removed in the flow-through step. In the first elution, the 140 kDa and 100 kDa peptides were detected, as was the 40 kDa band, which did not appear in the immunomagnetic separation without alkaline phosphatase treatment (fig. 6). This difference can be caused by the dissociation of the 40 kDa MCEF peptide from the complex, possibly mediated by the affinity of the alkaline phosphatase enzyme for the 40 kDa peptide. The second elution contains less of the full-length MCEF protein and hardly

contains any 100 kDa MCEF. This result is expected because most of the MCEF entities in the nuclear extract would have already been pulled out in the first elution; only peptides that tightly bound would be washed out in the second elution.

4.4 MCEF Deletion Analysis

The MCEF deletion analysis was conducted to identify the regions of MCEF that are thought to be responsible for down-regulating and repressing the Tat-transactivated transcription of HIV-1. In other words, the objective was to map the repression domains of the MCEF protein, an effort that can provide insight on both the function and the structure details of MCEF and pave the way for the implementation of HIV-1 chimeric transcription factor repression (CTFR) (see section 5 below).

The luciferase assay vectors have been constructed to specifically report the binding activity of a single transcription factor. Once transfected, the cells will express MCEF and Tat independently of the LTR plasmid. As mentioned previously, MCEF is thought to specifically repress the Tat-transactivation of HIV-1 replication through its interaction with P-TEFb. Theoretically, since both MCEF and Tat are expressed in the cells, the measure of HIV-1 transcriptional activity will depend entirely on the activity of the promoter. The luciferase enzyme will only be produced if the promoter region in that plasmid is active. A strongly active promoter will produce large amounts of *Renilla* luciferase mRNA, whereas a weak or inactive promoter will produce little or no mRNA; the stronger the promoter the more luciferase mRNA will be produced. Any mRNA produced will be translated into active luciferase enzyme that generates a signal that is used to detect transcriptional activity of the promoter. The strength of the promoter activity will depend on the activation by Tat and the subsequent repression by MCEF. Using different sequences of MCEF to be transfected will help in identifying which

region of MCEF will most efficiently suppress the Tat-transactivation of HIV-1.

The results in figure 8 demonstrate that the lowest level of luciferase expression was reported in the cells that expressed amino acids 694-1163 of MCEF, i.e. the C-terminal region of the protein. This observation suggests that the promoter activity in these cells was the lowest, which in turn suggests that the C-terminal of MCEF has the highest HIV-1 replication repression potential, which indicates that the repression domain of MCEF could possibly be located in the region between amino acids 694-1163. The difference in luciferase activity between this construct and the 1-715 construct does not seem significant, although the latter results in the production of slightly more luciferase enzyme (and therefore more active promoter). It is also important to note here the difference in luciferase expression between the 1-715 reaction and the 402-745 reaction. The amino acids 402-715 overlap in the two constructs, but the former contains the N-terminal region of the MCEF protein, while the latter does not. This suggests that the difference in transcriptional activity can be attributed to the activity of the N-terminal region of the protein. By considering the small difference in 1-715 and 694-1163, as well as the considerable difference in 1-715 and 402-745 (in terms of their effect on the expression of luciferase) together, it may be proposed that the N-terminal 1-402 region of MCEF could also repress Tat-transactivation. The +pcMCEF-transfected cells did not demonstrate any repression of HIV-1 Tat transactivation. The level of luciferase expression was reasonably high, and that can be explained by the fact that functional MCEF protein was not expressed, rather a protein with a reverse MCEF sequence was. Reversing the sequence of a protein essentially means that the normal functional protein can no longer be produced.

4.5 RBEIII Mutation Analysis

As mentioned in section 1.4, the RBEIII site in the HIV-1 LTR is a target for activating HIV-1 through the Ras-signaling pathway, putatively through TFIIi and USF actions at this site, although it is not entirely known whether this occurs due to binding or due to the phosphorylation and activation of an already bound TFIIi. A specific region in the RBEIII site, with the sequence ACTGCTGACATC has been found to be conserved in 100% of HIV-1 LTRs sampled in vivo, in HIV-1 infected persons [47, 48, 97]. This CTG duplication motif is conserved in HIV-1 LTRs because a compensatory mechanism duplicates the un-mutated RBEIII region [47, 97, 98]. This mechanism is mediated by a length polymorphism, known as the most frequent naturally occurring length polymorphism (MFNLP) [47, 97, 98].

The site directed mutagenesis experiment was conducted to study the effect of mutations in the CTG motif on the replication of live HIV-1. The CTG-motif has been previously studied by several groups [99, 100] and attempts at elucidating its significance have been recently reviewed [97]. As discovered by Dr. Estable, Mojgan Naghavi and Robert Roeder [50, 97, 101], RBF-2 consists of USF and TFIIi [50, 97, 101]. More recently, base pairs required for USF and TFIIi activity, have been suggested [90]. Although mutation of specific base pairs impaired Ras-responsiveness, it did not ablate transcription in their study [90].

In our experiments, two T-cell stimulants, CD3 and PHA, were used to mimic T-cell receptor signaling in the cells (as per Malcolm *et al.*), and two of the mutant HIV-1 constructs used by Sadowski *et al.* (mut-G and mut-M, referred to M4 and M1 here) were generated and for the first time tested *in vivo*, in live replication assays. As mentioned previously, the levels of p24 are often used as an indicator of HIV-1 replication because it is a protein that is incorporated in the capsid of the virus and its utilization indicates the production of new HIV-1 particles.

The mock wells were used as a negative control, since they did not contain HIV-1 DNA and therefore did not yield any HIV-1, as reflected by the absence of p24 production, at 48 hours and 96 hours following treatment. As expected, no p24 was detected in the "mock" under any conditions, indicating little to no background in the assay (see fig. 9). Without treatment, the positive control (monomer) construct produced double the amount of p24 after 96 hours compared to 48 hours. This suggests that USF/TFIIi has successfully interacted with the wild-type RBEIII site in response to normal levels of T-cell signaling through T-cell receptors. The levels of p24 produced in the PHA treated cells were double the levels of p24 in untreated cells, at 48 hours. This suggests that stimulation resulted in activation of the HIV-1 LTR, resulting in twice the level of HIV-1 replication. Between 48 hours and 96 hours following PHA treatment, the expression of p24 increased 3.5-fold. At 48 hours asynchronous cells would not have all doubled, so the activation is likely due to LTR-activation, rather than proliferation. However, the increase between 48 and 96 hours could be due to both cell proliferation and HIV-1 LTR activation.

The expression of p24 in the first 48 hours post anti-CD3 treatment was the same as that in the first 48 hours following PHA treatment, but this level remained constant until 96 hours post treatment, contrary to the results discussed for PHA. The pattern of responsiveness of T-cells to anti-CD3 treatment is independent of the number of cells growing in a sample. At the 48-hour and 96-hour time points the same number of viable cells was present, expressing the same amount of p24. Although this result suggests proliferation as the major component of the previous result, further experiments would be required to dissect-out the exact events responsible for this.

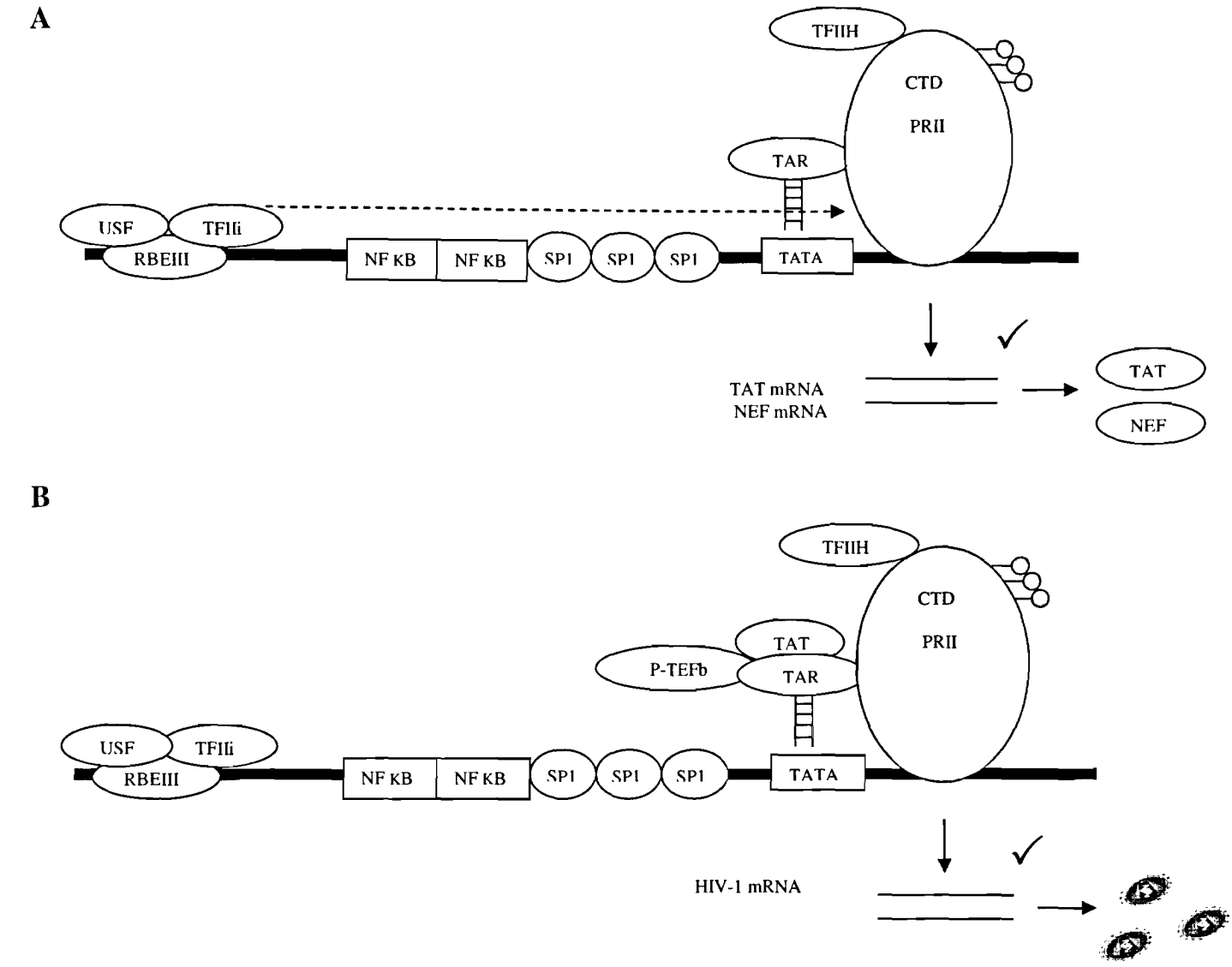
Importantly and unexpectedly, PHA and anti-CD3 treatments did not cause any

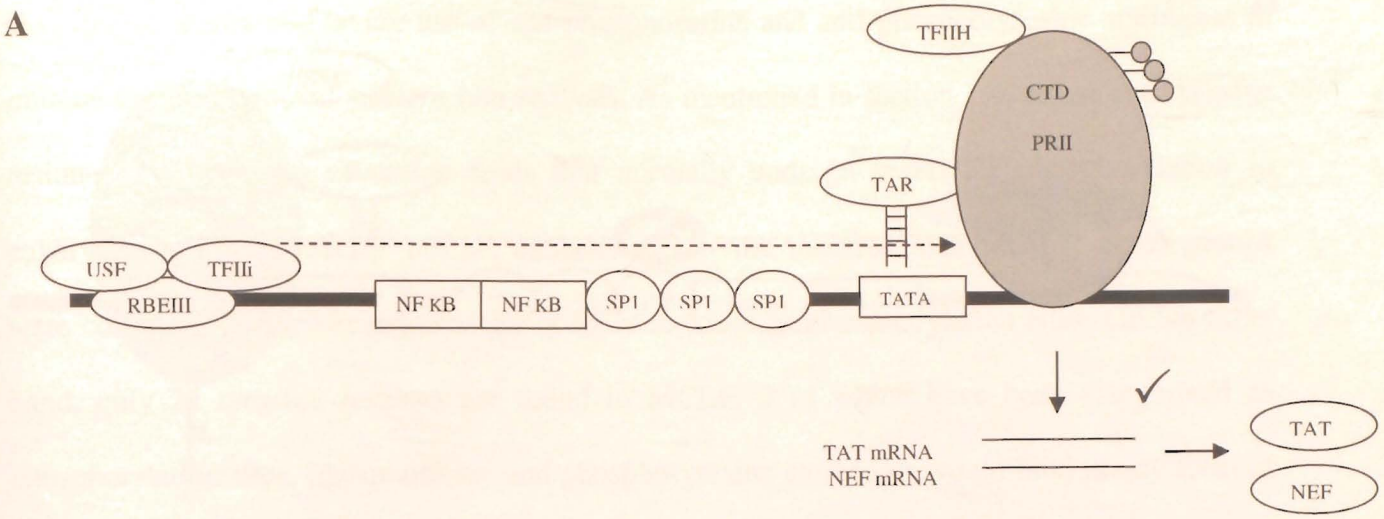
proliferation in the M1 and M4 constructs; in fact, no p24 was detected in either construct at any of the time points. This observation could mean that the mutations in the M1 and M4 constructs were able to completely repress the replication of HIV-1 in the cells. This result is contrary to that obtained by Sadowski *et al.*, where these mutations appeared to repress only the activation effects of the Ras pathway [90]. One possibility is that contrary to the previous experiments by Sadowski, our results used full-length replication-competent infectious clones, as compared to his groups limited replication assays using a luciferase reporter [90]. Alternatively, the lack of HIV-1 replication from the mutant constructs could also be a consequence of the LTR acquiring additional mutations (other than the 1 bp and 4 bp intended mutations) during the site directed mutagenesis reaction. Therefore, it is possible that together the 1 bp and 4bp mutations and the acquired mutations caused this significant repression in HIV-1 replication. More experiments will be required to sort this out.

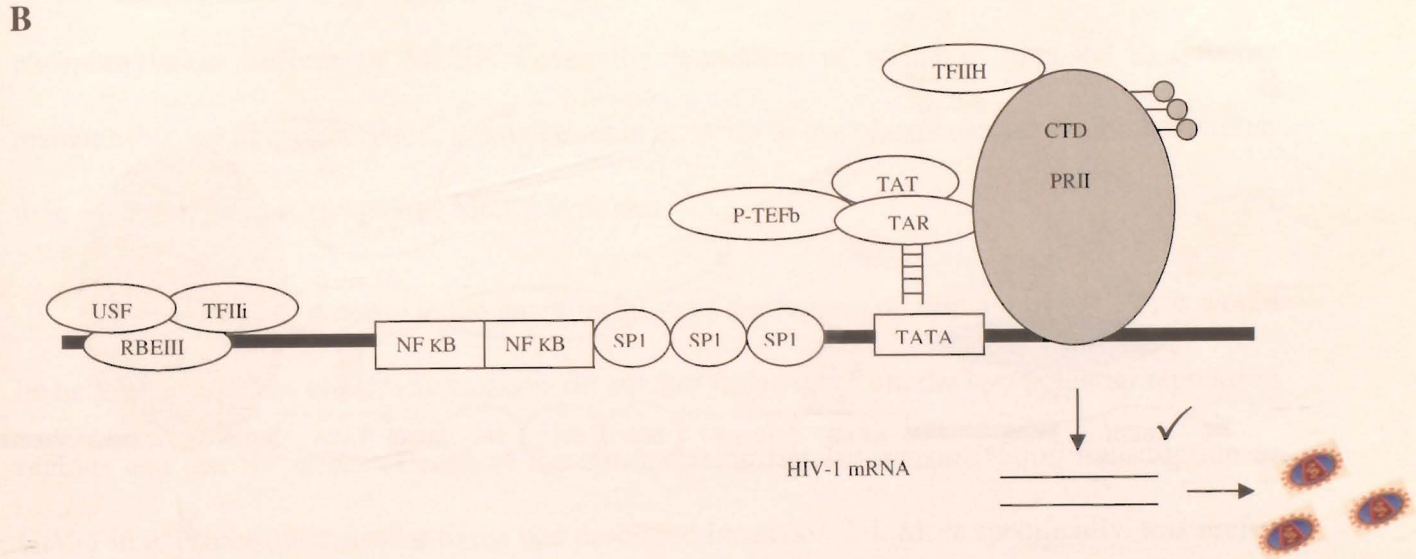
5. RECOMMENDATIONS

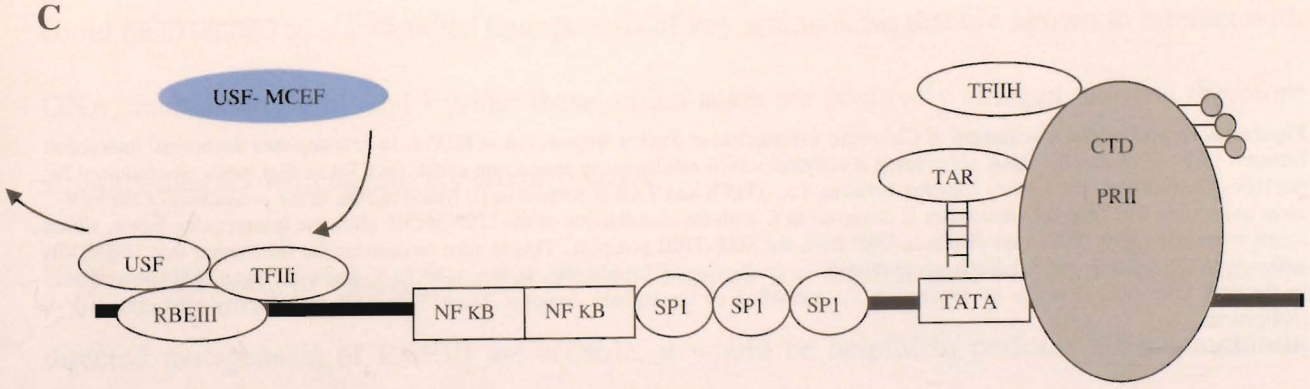
The previous discussion has demonstrated that not only are the roles of MCEF in need of further research, but they are also very diverse. The Tat-transactivation of HIV-1 has been shown to be significantly inhibited when specific sequences of the MCEF protein are deleted, supposedly due to its compromised ability to interact with the Cdk-9 subunit of P-TEFb. In order to further investigate this repressive ability of MCEF, it has been suggested that Chimeric Transcription Factor Repression (CTFR) can be explored [51]. Although the HIV-1 LTR is constantly modified, there is one region at the enhancer region of the HIV-1 LTR that is highly conserved, the RBEIII/(MFNLP) (see section 4.5). The RBF-2 factor consists of USF and TFIIi subunits that interact with the RBEIII/MFNLP site. If a chimeric protein, that consists partly of USF and partly of MCEF (the region that repressed Tat-transactivation) (USF-MCEF), is generated, then it may be possible to deliver MCEF to the LTR and amplify its' repressive activity on Tat-transactivation, as well as Ras-activation. Importantly, the binding to the conserved RBEIII region, would emulate the M1 and M4 mutations, described here.

The concept of engineered chimeric factors has been tested previously in HIV-1 studies and found to be successful [102]. In their study, Reynolds *et al.* were able to demonstrate that chimeric zing-finger proteins that were tethered to the HIV-1 LTR via an effector domain repressed HIV-1 Tat-transactivation more effectively than chimeric proteins that were un-tethered. The Estable laboratory hypothetical mechanism of RBF-2/MCEF CTFR is depicted in figure 10 and is outlined in a recent paper [97].


A
TFIIH
CTD
PRII
TAR
USF
TFIIi
RBEIII
NF κB
NF κB
SP1
SP1
SP1
TATA
TAT mRNA
NEF mRNA
TAT
NEF
B
TFIIH
CTD
PRII
TAT
P-TEFb
TAR
USF
TFIIi
RBEIII
NF κB
NF κB
SP1
SP1
SP1
TATA
HIV-1 mRNA
C
USF- MCEF
TFIIH
CTD
PRII
TAR
USF
TFIIi
RBEIII
NF κB
NF κB
SP1
SP1
SP1
TATA


A
TFIIH
CTD
PRII
TAR
USF
TFIIi
RBEIII
NF κB
NF κB
SP1
SP1
SP1
TATA
TAT mRNA
NEF mRNA
TAT
NEF


B
TFIIH
CTD
PRII
TAT
P-TEFb
TAR
USF
TFIIi
RBEIII
NF κB
NF κB
SP1
SP1
SP1
TATA
HIV-1 mRNA


C
USF- MCEF
TFIIH
CTD
PRII
TAR
USF
TFIIi
RBEIII
NF κB
NF κB
SP1
SP1
SP1
TATA

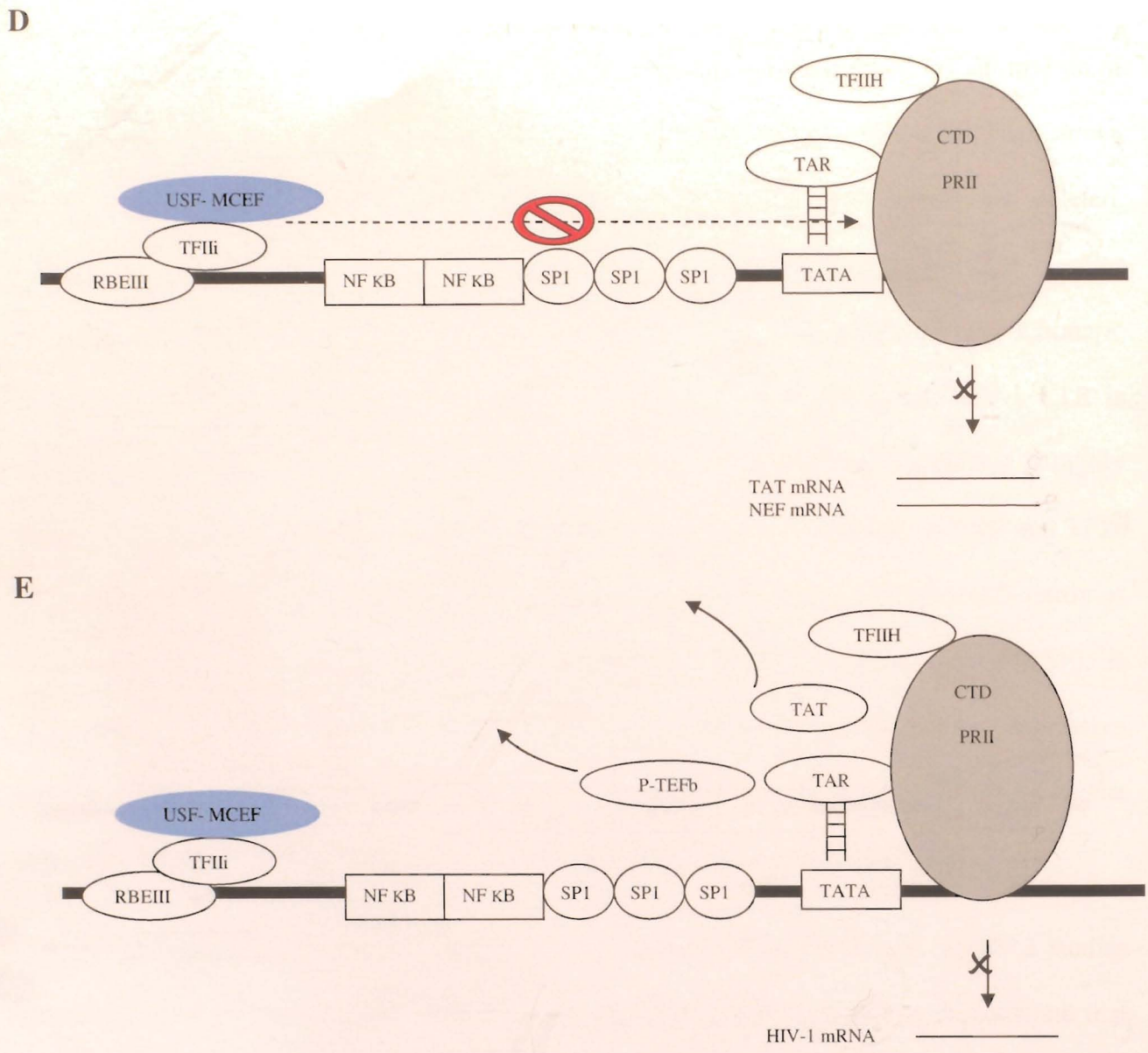


Figure 10. Hypothesized mechanism of Chimeric Transcription Factor Repression of HIV-1. In A is depicted the normal interaction between USF- TFIIi and the RNA polymerase II enzyme, which results in the production of the first Tat an Nef molecules required for the HIV-1 replication. The ternary complex between Tat, pTEFb and TAR is formed in B, which results in the production of the HIV-1 virus molecules. This normal mechanism is disrupted in C with the introduction of the USF-MCEF chimeric transcription factor, which forms a complex with TFIIi that displaces USF from the USF-TFIIi complex. This in turn terminates the interaction that USF-TFIIi makes with RNA pol II, which completely abolishes the production of Tat and Nef, as seen in D. In E, when Tat and pTEFb are available in the cell, USF-MCEF works by preventing the formation of the ternary complex Tat-pTEFb-TAR and thereby inhibiting HIV-1 replication.

Recommendations can also be made for the studies already outlined in the paper. For example, to be able to make conclusions about phosphorylation isoforms of MCEF, a reasonable

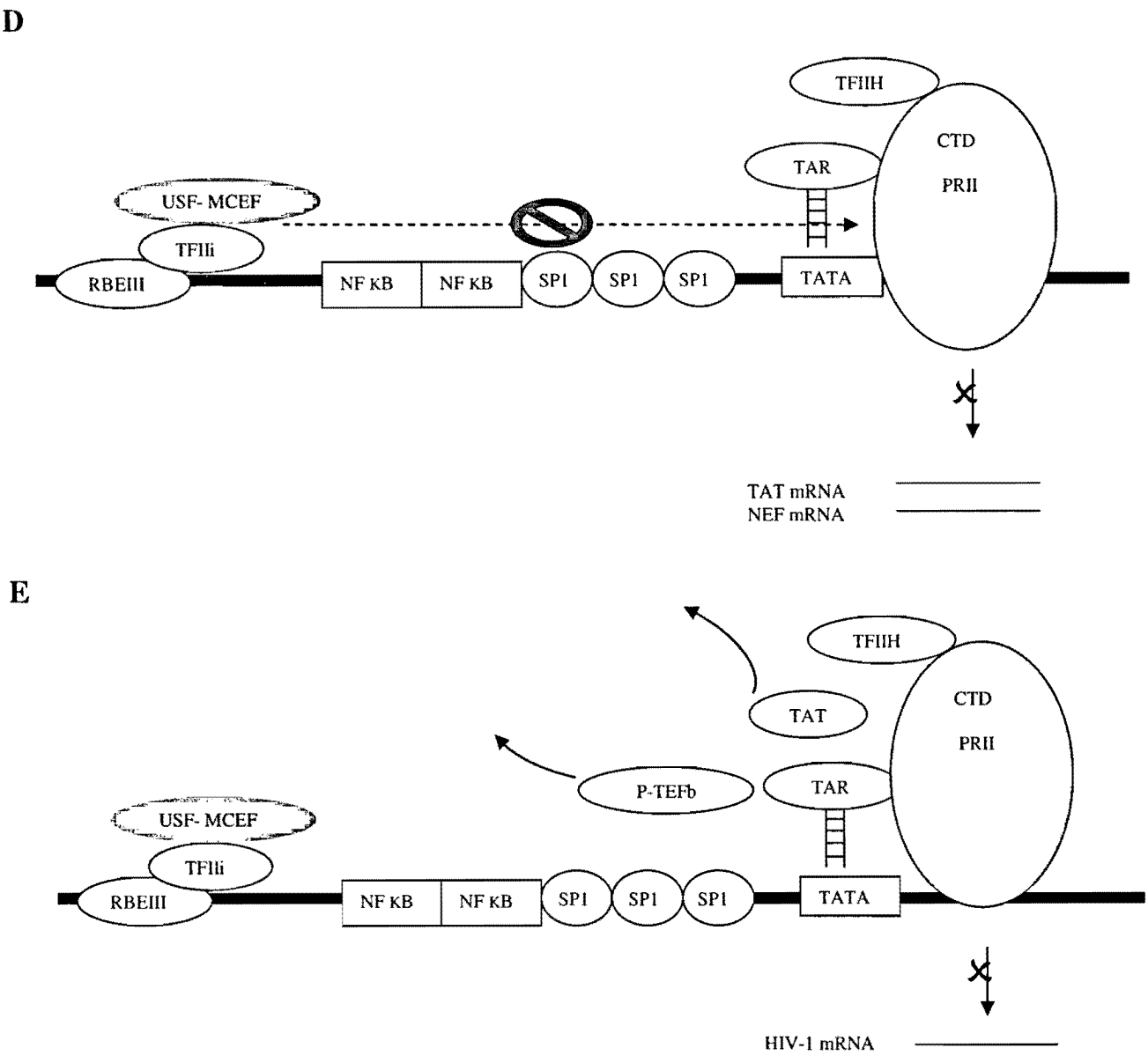


Figure 10. Hypothesized mechanism of Chimeric Transcription Factor Repression of HIV-1. In A is depicted the normal interaction between USF- TFIIi and the RNA polymerase II enzyme, which results in the production of the first Tat an Nef molecules required for the HIV-1 replication. The ternary complex between Tat, pTEFb and TAR is formed in B, which results in the production of the HIV-1 virus molecules. This normal mechanism is disrupted in C with the introduction of the USF-MCEF chimeric transcription factor, which forms a complex with TFIIi that displaces USF from the USF-TFIIi complex. This in turn terminates the interaction that USF-TFIIi makes with RNA pol II, which completely abolishes the production of Tat and Nef, as seen in D. In E, when Tat and pTEFb are available in the cell, USF-MCEF works by preventing the formation of the ternary complex Tat-pTEFb-TAR and thereby inhibiting HIV-1 replication.

Recommendations can also be made for the studies already outlined in the paper. For example, to be able to make conclusions about phosphorylation isoforms of MCEF, a reasonable

next step to take would be the use of anti-phosphoserine and anti-phosphotyrosine antibodies in immunoprecipitation and western blot analysis. As mentioned in section 1.4, serine and tyrosine residues are examples of amino acids that normally undergo reversible phosphorylation in eukaryotic cells. The MCEF protein contains 129 serine residues; individual research groups were collectively able to categorize 39 of these residues as phosphorylation sites. On the other hand, only 26 tyrosine residues are found in MCEF, 3 of which have been categorized as phosphorylation sites. Phosphoserine and phosphotyrosine antibodies would bind to any form of MCEF that contains phosphorylated serine and/or tyrosine residues to identify the phosphorylation isoform of MCEF. Given the proportion of serine as opposed to tyrosine residues that are phosphorylated, it may be more accurate to use phosphoserine antibodies, rather than phosphotyrosine, to identify MCEF isoforms.

In addition, to determine the exact location of the repression domain of MCEF, it would be helpful to generate constructs that contain smaller fragments from the two potential repression regions and test the effect of each of the constructs on the Tat-transactivation transcription of HIV-1 in an experiment similar to the one described in section 2.4. More specifically, this notion could be extended to site directed mutagenesis of key amino acids that are known to interact with DNA, such as Arginine and Lysine; these amino acids are positively charged and can therefore strongly interact with negatively charged DNA.

Moreover, to ensure that the results obtained for the M1 and M4 mutants in the site directed mutagenesis of RBEIII are reliable, it would be helpful to perform a back mutation experiment. Essentially, the sequences of the two mutants would be returned to the original sequence found in the wild-type form of the CTG motif and analyzed again for their effect on HIV-1 replication; if replication was restored, then the results can be considered conclusive and

the findings can be used as a platform for therapeutic studies. Sequencing and back-mutation analysis would also determine whether it was just the M1 and M4 mutations that resulted in HIV-1 replication repression and not acquired mutations (if any).

6. CONCLUSION

The role of humans in the deteriorating state of the environment may have been established, but the effect of unsuitable chemical and biological environmental factors on fetus health remain unclear, although some of these have been identified as risk factors. The purpose of this paper was to provide insight on the research that is currently being done on the characterization of the MCEF protein, whose modification (presumably due to toxic environmental conditions in the uterus) leads to acute lymphoblastic leukemia. MCEF has been identified as the fourth member of the AF4 family of transcription factors.

A phylogenetic analysis was carried out to study the evolutionary descents of the AF4 proteins and the results indicated that AFF2 and FF3 formed sister clades, while AFF1 and AFF4 did not. The second divergence from their common ancestor gave rise to the AFF4 protein. This suggests that the AF4 proteins developed from a single protein sequence by ancestral duplication. A detailed examination of the results of western blots conducted with different cellular lysates and different primary anti-MCEF antibodies was also provided and the results suggested that MCEF may have isoforms with molecular weights of 100 kDa and 40 kDa, produced by alternative splicing of the MCEF mRNA. The results also suggested that MCEF may be post-translationally modified by phosphorylation. Accordingly, experiments were conducted to determine whether or not this is true, and although no conclusive results were obtained, there is no doubt that the activity of the 40 kDa peptide appears to be tightly regulated.

Some evidence has suggested that MCEF could also play a role in the repression of the Tat-transactivation of HIV-1 transcription through its interaction with P-TEFb. It is thought that this interaction compromises the ability of P-TEFb to form a ternary complex with Tat and TAR that in necessary if HIV-1 transcription were to occur. In order to map the regions of MCEF that

could possibly be interacting with P-TEFb and repressing transcription, the effect of different MCEF deletion mutants on the transcription of HIV-1 was tested. The results suggested that the region between 1- 402, as well as the C terminal region, were most efficiently able to down-regulate HIV-1 transcription, suggesting the possibility of MCEF containing two repression domains, rather than just one. In relation to the research on HIV-1, the CTG motif near the RBEIII site in the HIV-1 LTR was also mutated in a site directed mutagenesis experiment to determine whether this motif (and the proteins that it binds) can be used to repress HIV-1 replication. The results showed that the mutant CTG constructs were able to completely inhibit the replication of HIV-1, which indicates that the mutations significantly impaired the capacity of RBEIII to respond to the activation of the Ras-pathway.

It is evident from the research discussed previously that the role of MCEF is still unknown and that there is plenty of progress to be done in various areas. It is also evident that MCEF is involved in a diverse range of cellular pathways that could lead to disease if their regulation was lost due to the availability of suitable environmental conditions. The importance here lies in the elucidation of the structural and functional relevance of MCEF, bringing us closer towards the treatment of patients suffering from diseases associated with the protein, and the eventual betterment and advancement of human health and their quality of life.

7. APPENDIX

Table A-1. Proposed phosphorylation sites of human MCEF. Information obtained from GenBank and Phosphosite Plus. S: Serine; K: Threonine; Y: Tyrosine. No. of refs refers to the number of authors that have reported the residue as a phosphorylation site for MCEF. Phosphorylated residues are in red.

Residue	Sequence	Residue	Sequence
S31	GEDAFPPssPLFAEP	S514	REQGTGNsYTDTSGP
S32	EDAFPPssPLFAEPY	T528	PKETSSAtPGRDSKT
K79	PKLVAIPkPTVPPSA	S549	SGRGRQKsPAQSDST
S124	TSQSQKRsSGLQSGH	S599	LASSMPSsRHKAATK
S176	HGSEHSKsRsssPGK	S617	KPNIKKEsKSSPRPT
S178	SEHSKsRsssPGKPQ	S668	PSSQTPKyPEsNRtP
S179	EHSKsRsssPGKPQA	S671	QTPKyPEsNRtPVKP
S180	HSKsRsssPGKPQAV	T674	KyPEsNRtPVKPssV
S193	AVSSLNSsHsRSHGN	S679	NRtPVKPssVEEEDS
S195	SSLNSsHsRSHGNDH	S680	RtPVKPssVEEEDSF
S210	HSKEHQRsKsPRDPD	S694	FFRQRMFsPMEEKEL
S212	KEHQRsKsPRDPDAN	S703	MEEKELLsPLsEPDD
S222	DPDANWDsPSRVPFS	S706	KELLsPLsEPDDRyP
S387	LKDDLKLsssEDsDG	Y712	LsEPDDRyPLIVKID
S388	KDDLKLsssEDsDGE	S814	KEKDLLPsPAGPVPs
S389	DDLKLsssEDsDGEQ	S821	sPAGPVPsKDPKTEH
S392	KLsssEDsDGEQDCD	S836	GSRKRTIsQSSSLKS
S487	KVNPHKVsPAssVDS	S1039	LKNSYNNsQAPsPGL
S490	PHKVsPAssVDSNIP	S1043	YNNsQAPsPGLGSKA
S491	HKVsPAssVDSNIPS	S1055	SKAVGMPsPVsPKLs
S499	VDSNIPSsQGyKKEG	S1058	VGMPsPVsPKLsPGN
Y502	NIPSsQGyKKEGREQ	S1062	sPVsPKLsPGNsGNY
		S1066	PKLsPGNsGNYSSGA

Table A-2. Luciferase data for the MCEF deletion experiment.

	Trial 1	Trial 2	Trial 3	Average
1- 715	5964558	7286978	6870383	6707306
402- 745	20207407	18844261	19747843	19599837
694- 1163	4950611	5450916	4630787	5010771
+ pcMCEF	19485590	18715563	16225649	18142267

Table A-3. p24 absorbance values for the RBEIII mutagenesis experiment.

a. 48 hours

	No Treatment			PHA (10 µg/mL)			anti-CD3 (2.5 µg/mL)		
	Trial 1	*Trial 2*	*Trial 3*	*Trial 1*	*Trial 2*	*Trial 3*	*Trial 1*	*Trial 2*	*Trial 3*
Mock	0.007	0.008	0.007	0.007	0.009	-0.005	0.005	0.008	0.005
Monomer	0.149	0.182	0.191	0.234	0.324	0.341	0.274	0.247	0.299
M1	0.009	0.001	0.006	-0.002	0.003	-0.002	0.001	-0.006	0.002
M4	0.007	0.001	-0.004	-0.003	0.003	0	0.011	0.004	0.005

b. 96 hours

	No Treatment			PHA (10 µg/mL)			anti-CD3 (2.5 µg/mL)		
	Trial 1	*Trial 2*	*Trial 3*	*Trial 1*	*Trial 2*	*Trial 3*	*Trial 1*	*Trial 2*	*Trial 3*
Mock	-0.004	-0.004	0	-0.006	-0.001	0.001	0.005	-0.003	-0.005
Monomer	0.239	0.333	0.313	0.661	1.171	1.339	0.228	0.218	0.369
M1	0.001	0.001	0.002	0.001	0	0.001	0	0.001	0.018
M4	-0.001	-0.001	-0.002	0.006	0.001	0	0.003	0.005	0.001

Table A-2. Luciferase data for the MCEF deletion experiment.

	Trial 1	Trial 2	Trial 3	Average
1- 715	5964558	7286978	6870383	6707306
402- 745	20207407	18844261	19747843	19599837
694- 1163	4950611	5450916	4630787	5010771
+ pcMCEF	19485590	18715563	16225649	18142267

Table A-3. p24 absorbance values for the RBEIII mutagenesis experiment.

a. 48 hours

	No Treatment			PHA (10 µg/mL)			anti-CD3 (2.5 µg/mL)		
	Trial 1	*Trial 2*	*Trial 3*	*Trial 1*	*Trial 2*	*Trial 3*	*Trial 1*	*Trial 2*	*Trial 3*
Mock	0.007	0.008	0.007	0.007	0.009	-0.005	0.005	0.008	0.005
Monomer	0.149	0.182	0.191	0.234	0.324	0.341	0.274	0.247	0.299
M1	0.009	0.001	0.006	-0.002	0.003	-0.002	0.001	-0.006	0.002
M4	0.007	0.001	-0.004	-0.003	0.003	0	0.011	0.004	0.005

b. 96 hours

	No Treatment			PHA (10 µg/mL)			anti-CD3 (2.5 µg/mL)		
	Trial 1	*Trial 2*	*Trial 3*	*Trial 1*	*Trial 2*	*Trial 3*	*Trial 1*	*Trial 2*	*Trial 3*
Mock	-0.004	-0.004	0	-0.006	-0.001	0.001	0.005	-0.003	-0.005
Monomer	0.239	0.333	0.313	0.661	1.171	1.339	0.228	0.218	0.369
M1	0.001	0.001	0.002	0.001	0	0.001	0	0.001	0.018
M4	-0.001	-0.001	-0.002	0.006	0.001	0	0.003	0.005	0.001

8. REFERENCES

1. Muthayya, S., *Maternal nutrition & low birth weight - what is really important?* Indian J Med Res, 2009. **130**(5): p. 600-8.
2. Hall, J. and F. Solehdin, *Folic acid for the prevention of congenital anomalies.* Eur J Pediatr, 1998. **157**(6): p. 445-50.
3. Li, Y.F., et al., *Effects of in utero and environmental tobacco smoke exposure on lung function in boys and girls with and without asthma.* Am J Respir Crit Care Med, 2000. **162**(6): p. 2097-104.
4. Meador, K.J., et al., *Cognitive function at 3 years of age after fetal exposure to antiepileptic drugs.* N Engl J Med, 2009. **360**(16): p. 1597-605.
5. Ron, E., *Ionizing radiation and cancer risk: evidence from epidemiology.* Pediatr Radiol, 2002. **32**(4): p. 232-7; discussion 242-4.
6. Holick, M.F., *Vitamin D deficiency.* N Engl J Med, 2007. **357**(3): p. 266-81.
7. McKay, J.A., E.A. Williams, and J.C. Mathers, *Folate and DNA methylation during in utero development and aging.* Biochem Soc Trans, 2004. **32**(Pt 6): p. 1006-7.
8. Cooper, C., et al., *Growth in infancy and bone mass in later life.* Ann Rheum Dis, 1997. **56**(1): p. 17-21.
9. Hales, C.N. and D.J. Barker, *Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis.* Diabetologia, 1992. **35**(7): p. 595-601.
10. Kensara, O.A., et al., *Fetal programming of body composition: relation between birth weight and body composition measured with dual-energy X-ray absorptiometry and anthropometric methods in older Englishmen.* Am J Clin Nutr, 2005. **82**(5): p. 980-7.
11. Osmond, C., et al., *Early growth and death from cardiovascular disease in women.* BMJ, 1993. **307**(6918): p. 1519-24.
12. Lucas, A., *Programming by early nutrition in man.* Ciba Found Symp, 1991. **156**: p. 38-50; discussion 50-5.
13. Buka, I., S. Koranteng, and A.R. Osornio Vargas, *Trends in childhood cancer incidence: review of environmental linkages.* Pediatr Clin North Am, 2007. **54**(1): p. 177-203, x.
14. Draper, G.J., et al., *Cancer in the offspring of radiation workers: a record linkage study.* BMJ, 1997. **315**(7117): p. 1181-8.
15. McHale, C.M. and M.T. Smith, *Prenatal origin of chromosomal translocations in acute childhood leukemia: implications and future directions.* Am J Hematol, 2004. **75**(4): p. 254-7.
16. Wiemels, J.L., et al., *In utero origin of t(8;21) AML1-ETO translocations in childhood acute myeloid leukemia.* Blood, 2002. **99**(10): p. 3801-5.
17. Torpy, J.M., C. Lynm, and R.M. Glass, *JAMA patient page. Acute lymphoblastic leukemia.* JAMA, 2009. **301**(4): p. 452.
18. Pui, C.H., L.L. Robison, and A.T. Look, *Acute lymphoblastic leukaemia.* Lancet, 2008. **371**(9617): p. 1030-43.
19. Sadananda Adiga, M.N., et al., *Homocysteine, vitamin B12 and folate status in pediatric acute lymphoblastic leukemia.* Indian J Pediatr, 2008. **75**(3): p. 235-8.
20. Yang, W., et al., *ARID5B SNP rs10821936 is associated with risk of childhood acute lymphoblastic leukemia in blacks and contributes to racial differences in leukemia incidence.* Leukemia, 2010. **24**(4): p. 894-896.
21. Semsei, A.F., P. Antal, and C. Szalai, *Strengths and weaknesses of gene association studies in childhood acute lymphoblastic leukemia.* Leuk Res, 2010. **34**(3): p. 269-71.

22. Wiemels, J.L., et al., *Methylenetetrahydrofolate reductase (MTHFR) polymorphisms and risk of molecularly defined subtypes of childhood acute leukemia.* Proc Natl Acad Sci U S A, 2001. **98**(7): p. 4004-9.
23. Hengstler, J.G., Heimerdinger, C.K., Schiffer, I.B., Gebhard, S., Sagemuller, J., Tanner, B., Bolt, H.M., Oesch, F., *Dietary topoisomerase II-poisons: contribution of soy products to infant leukemia.* EXCLI, 2002. **1**: p. 8.
24. Ayton, P.M. and M.L. Cleary, *Molecular mechanisms of leukemogenesis mediated by MLL fusion proteins.* Oncogene, 2001. **20**(40): p. 5695-707.
25. Meyer, C., et al., *New insights to the MLL recombinome of acute leukemias.* Leukemia, 2009. **23**(8): p. 1490-9.
26. Meyer, C., et al., *Diagnostic tool for the identification of MLL rearrangements including unknown partner genes.* Proc Natl Acad Sci U S A, 2005. **102**(2): p. 449-54.
27. Armstrong, S.A., et al., *MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia.* Nat Genet, 2002. **30**(1): p. 41-7.
28. Taki, T., et al., *AF5q31, a newly identified AF4-related gene, is fused to MLL in infant acute lymphoblastic leukemia with ins(5;11)(q31;q13q23).* Proc Natl Acad Sci U S A, 1999. **96**(25): p. 14535-40.
29. Deveney, R., et al., *Insertion of MLL sequences into chromosome band 5q31 results in an MLL-AF5Q31 fusion and is a rare but recurrent abnormality associated with infant leukemia.* Genes Chromosomes Cancer, 2003. **37**(3): p. 326-31.
30. Bursen, A., et al., *The AF4.MLL fusion protein is capable of inducing ALL in mice without requirement of MLL.AF4.* Blood, 2010. **115**(17): p. 3570-9.
31. Chen, C.S., et al., *The chromosome 4q21 gene (AF-4/FEL) is widely expressed in normal tissues and shows breakpoint diversity in t(4;11)(q21;q23) acute leukemia.* Blood, 1993. **82**(4): p. 1080-5.
32. Chen, W., et al., *A murine Mll-AF4 knock-in model results in lymphoid and myeloid deregulation and hematologic malignancy.* Blood, 2006. **108**(2): p. 669-77.
33. Imamura, T., et al., *A novel infant acute lymphoblastic leukemia cell line with MLL-AF5q31 fusion transcript.* Leukemia, 2002. **16**(11): p. 2302-8.
34. Kaatsch, P., *Epidemiology of childhood cancer.* Cancer Treat Rev, 2010. **36**(4): p. 277-285.
35. Thirman, M.J., et al., *Rearrangement of the MLL gene in acute lymphoblastic and acute myeloid leukemias with 11q23 chromosomal translocations.* N Engl J Med, 1993. **329**(13): p. 909-14.
36. Zhu, Y., et al., *Transcription elongation factor P-TEFb is required for HIV-1 tat transactivation in vitro.* Genes Dev, 1997. **11**(20): p. 2622-32.
37. Wei, P., Garber, M.E., Fang, S.M., Fischer, W.H., Jones, K.A. , *A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA.* Cell, 1998. **92**: p. 451.
38. Estable, M.C., et al., *MCEF, the newest member of the AF4 family of transcription factors involved in leukemia, is a positive transcription elongation factor-b-associated protein.* J Biomed Sci, 2002. **9**(3): p. 234-45.
39. Peng, J., N.F. Marshall, and D.H. Price, *Identification of a cyclin subunit required for the function of Drosophila P-TEFb.* J Biol Chem, 1998. **273**(22): p. 13855-60.

40. Chen, D. and Q. Zhou, *Tat activates human immunodeficiency virus type 1 transcriptional elongation independent of TFIIH kinase.* Mol Cell Biol, 1999. **19**(4): p. 2863-71.
41. Isel, C. and J. Karn, *Direct evidence that HIV-1 Tat stimulates RNA polymerase II carboxyl-terminal domain hyperphosphorylation during transcriptional elongation.* J Mol Biol, 1999. **290**(5): p. 929-41.
42. Ping, Y.H. and T.M. Rana, *Tat-associated kinase (P-TEFb): a component of transcription preinitiation and elongation complexes.* J Biol Chem, 1999. **274**(11): p. 7399-404.
43. Price, D.H., *P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II.* Mol Cell Biol, 2000. **20**(8): p. 2629-34.
44. Qiu, H., C. Hu, and A.G. Hinnebusch, *Phosphorylation of the Pol II CTD by KIN28 enhances BUR1/BUR2 recruitment and Ser2 CTD phosphorylation near promoters.* Mol Cell, 2009. **33**(6): p. 752-62.
45. Phatnani, H.P. and A.L. Greenleaf, *Phosphorylation and functions of the RNA polymerase II CTD.* Genes Dev, 2006. **20**(21): p. 2922-36.
46. Wimmer, J., et al., *Interactions between Tat and TAR and human immunodeficiency virus replication are facilitated by human cyclin T1 but not cyclins T2a or T2b.* Virology, 1999. **255**(1): p. 182-9.
47. Estable, M.C., et al., *Naturally occurring human immunodeficiency virus type 1 long terminal repeats have a frequently observed duplication that binds RBF-2 and represses transcription.* J Virol, 1998. **72**(8): p. 6465-74.
48. Estable, M.C., et al., *Human immunodeficiency virus type 1 long terminal repeat variants from 42 patients representing all stages of infection display a wide range of sequence polymorphism and transcription activity.* J Virol, 1996. **70**(6): p. 4053-62.
49. Estable, M.C., et al., *Purification of RBF-2, a transcription factor with specificity for the most conserved cis-element of naturally occurring HIV-1 LTRs.* J Biomed Sci, 1999. **6**(5): p. 320-32.
50. Naghavi, M.H., et al., *Upstream stimulating factor affects human immunodeficiency virus type 1 (HIV-1) long terminal repeat-directed transcription in a cell-specific manner, independently of the HIV-1 subtype and the core-negative regulatory element.* J Gen Virol, 2001. **82**(Pt 3): p. 547-59.
51. Niedzielski, M.F., et al., *MCEF is localized to the nucleus by protein sequences encoded within three distinct exons, where it represses HIV-1 Tat-transactivation of LTR-directed transcription.* Int J Biol Sci, 2007. **3**(4): p. 225-36.
52. Patarca, R., *Protein phosphorylation and dephosphorylation in physiologic and oncologic processes.* Crit Rev Oncog, 1996. **7**(5-6): p. 343-432.
53. Cantin, G.T., et al., *Combining protein-based IMAC, peptide-based IMAC, and MudPIT for efficient phosphoproteomic analysis.* J Proteome Res, 2008. **7**(3): p. 1346-51.
54. Daub, H., et al., *Kinase-selective enrichment enables quantitative phosphoproteomics of the kinome across the cell cycle.* Mol Cell, 2008. **31**(3): p. 438-48.
55. Matsuoka, S., et al., *ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage.* Science, 2007. **316**(5828): p. 1160-6.
56. Olsen, J.V., et al., *Global, in vivo, and site-specific phosphorylation dynamics in signaling networks.* Cell, 2006. **127**(3): p. 635-48.

57. Dephoure, N., et al., *A quantitative atlas of mitotic phosphorylation.* Proc Natl Acad Sci U S A, 2008. **105**(31): p. 10762-7.
58. Gecz, J., et al., *Gene structure and subcellular localization of FMR2, a member of a new family of putative transcription activators.* Genomics, 1997. **44**(2): p. 201-13.
59. Gu, Y., et al., *Identification of FMR2, a novel gene associated with the FRAXE CCG repeat and CpG island.* Nat Genet, 1996. **13**(1): p. 109-13.
60. Ma, C. and L.M. Staudt, *LAF-4 encodes a lymphoid nuclear protein with transactivation potential that is homologous to AF-4, the gene fused to MLL in t(4;11) leukemias.* Blood, 1996. **87**(2): p. 734-45.
61. Nilson, I., et al., *Exon/intron structure of the human AF-4 gene, a member of the AF-4/LAF-4/FMR-2 gene family coding for a nuclear protein with structural alterations in acute leukaemia.* Br J Haematol, 1997. **98**(1): p. 157-69.
62. Jans, D.A., C.K. Chan, and S. Huebner, *Signals mediating nuclear targeting and their regulation: application in drug delivery.* Med Res Rev, 1998. **18**(4): p. 189-223.
63. Erfurth, F., et al., *MLL fusion partners AF4 and AF9 interact at subnuclear foci.* Leukemia, 2004. **18**(1): p. 92-102.
64. Bitoun, E., P.L. Oliver, and K.E. Davies, *The mixed-lineage leukemia fusion partner AF4 stimulates RNA polymerase II transcriptional elongation and mediates coordinated chromatin remodeling.* Hum Mol Genet, 2007. **16**(1): p. 92-106.
65. Isaacs, A.M., et al., *A mutation in Af4 is predicted to cause cerebellar ataxia and cataracts in the robotic mouse.* J Neurosci, 2003. **23**(5): p. 1631-7.
66. Oliver, P.L., et al., *Mediation of Af4 protein function in the cerebellum by Siah proteins.* Proc Natl Acad Sci U S A, 2004. **101**(41): p. 14901-6.
67. Frestedt, J.L., J.M. Hilden, and J.H. Kersey, *AF4/FEL, a gene involved in infant leukemia: sequence variations, gene structure, and possible homology with a genomic sequence on 5q31.* DNA Cell Biol, 1996. **15**(8): p. 669-78.
68. Gu, Y., et al., *The t(4;11) chromosome translocation of human acute leukemias fuses the ALL-1 gene, related to Drosophila trithorax, to the AF-4 gene.* Cell, 1992. **71**(4): p. 701-8.
69. Nakamura, T., et al., *Genes on chromosomes 4, 9, and 19 involved in 11q23 abnormalities in acute leukemia share sequence homology and/or common motifs.* Proc Natl Acad Sci U S A, 1993. **90**(10): p. 4631-5.
70. Abrams, M.T., et al., *Cognitive, behavioral, and neuroanatomical assessment of two unrelated male children expressing FRAXE.* Am J Med Genet, 1997. **74**(1): p. 73-81.
71. Gecz, J., et al., *Identification of the gene FMR2, associated with FRAXE mental retardation.* Nat Genet, 1996. **13**(1): p. 105-8.
72. Bensaid, M., et al., *FRAXE-associated mental retardation protein (FMR2) is an RNA-binding protein with high affinity for G-quartet RNA forming structure.* Nucleic Acids Res, 2009. **37**(4): p. 1269-79.
73. Cross, S.H. and A.P. Bird, *CpG islands and genes.* Curr Opin Genet Dev, 1995. **5**(3): p. 309-14.
74. Chakrabarti, L., et al., *A candidate gene for mild mental handicap at the FRAXE fragile site.* Hum Mol Genet, 1996. **5**(2): p. 275-82.
75. Hillman, M.A. and J. Gecz, *Fragile XE-associated familial mental retardation protein 2 (FMR2) acts as a potent transcription activator.* J Hum Genet, 2001. **46**(5): p. 251-9.

76. von Bergh, A.R., et al., *LAF4, an AF4-related gene, is fused to MLL in infant acute lymphoblastic leukemia.* Genes Chromosomes Cancer, 2002. **35**(1): p. 92-6.
77. Hiwatari, M., et al., *Fusion of an AF4-related gene, LAF4, to MLL in childhood acute lymphoblastic leukemia with t(2;11)(q11;q23).* Oncogene, 2003. **22**(18): p. 2851-5.
78. Bruch, J., et al., *Occurrence of an MLL/LAF4 fusion gene caused by the insertion ins(11;2)(q23;q11.2q11.2) in an infant with acute lymphoblastic leukemia.* Genes Chromosomes Cancer, 2003. **37**(1): p. 106-9.
79. To, M.D., et al., *LAF-4 is aberrantly expressed in human breast cancer.* Int J Cancer, 2005. **115**(4): p. 568-74.
80. Steichen-Gersdorf, E., et al., *Triangular tibia with fibular aplasia associated with a microdeletion on 2q11.2 encompassing LAF4.* Clin Genet, 2008. **74**(6): p. 560-5.
81. Barton, A., et al., *Identification of AF4/FMR2 family, member 3 (AFF3) as a novel rheumatoid arthritis susceptibility locus and confirmation of two further pan-autoimmune susceptibility genes.* Hum Mol Genet, 2009.
82. Durkin, S.G. and T.W. Glover, *Chromosome fragile sites.* Annu Rev Genet, 2007. **41**: p. 169-92.
83. Burrow, A.A., et al., *Over half of breakpoints in gene pairs involved in cancer-specific recurrent translocations are mapped to human chromosomal fragile sites.* BMC Genomics, 2009. **10**: p. 59.
84. Strissel, P.L., et al., *An in vivo topoisomerase II cleavage site and a DNase I hypersensitive site colocalize near exon 9 in the MLL breakpoint cluster region.* Blood, 1998. **92**(10): p. 3793-803.
85. Lin, C., et al., *AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia.* Mol Cell, 2010. **37**(3): p. 429-37.
86. Urano, A., et al., *Infertility with defective spermiogenesis in mice lacking AF5q31, the target of chromosomal translocation in human infant leukemia.* Mol Cell Biol, 2005. **25**(15): p. 6834-45.
87. Wittwer, F., et al., *Lilliputian: an AF4/FMR2-related protein that controls cell identity and cell growth.* Development, 2001. **128**(5): p. 791-800.
88. Laemmli, U.K., *Cleavage of structural proteins during the assembly of the head of bacteriophage T4.* Nature, 1970. **227**(5259): p. 680-5.
89. Patras, G., G.G. Qiao, and D.H. Solomon, *On the mechanism of background silver staining during sodium dodecyl sulphate-polyacrylamide gel electrophoresis.* Electrophoresis, 1999. **20**(10): p. 2039-45.
90. Malcolm, T., et al., *Specific interaction of TFII-I with an upstream element on the HIV-1 LTR regulates induction of latent provirus.* FEBS Lett, 2008. **582**(28): p. 3903-8.
91. Dignam, J.D., R.M. Lebovitz, and R.G. Roeder, *Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei.* Nucleic Acids Res, 1983. **11**(5): p. 1475-89.
92. Ernst, D.N., et al., *Stimulation of murine T cell subsets with anti-CD3 antibody. Age-related defects in the expression of early activation molecules.* J Immunol, 1989. **142**(5): p. 1413-21.
93. Kay, J.E., *Mechanisms of T lymphocyte activation.* Immunol Lett, 1991. **29**(1-2): p. 51-4.
94. Bess, J.W., Jr., et al., *Tightly bound zinc in human immunodeficiency virus type 1, human T-cell leukemia virus type I, and other retroviruses.* J Virol, 1992. **66**(2): p. 840-7.

95. Layne, S.P., et al., *Factors underlying spontaneous inactivation and susceptibility to neutralization of human immunodeficiency virus.* Virology, 1992. **189**(2): p. 695-714.
96. Peterson, K.J., et al., *Estimating metazoan divergence times with a molecular clock.* Proc Natl Acad Sci U S A, 2004. **101**(17): p. 6536-41.
97. Estable, M.C., *In search of a function for the most frequent naturally-occurring length polymorphism (MFNLP) of the HIV-1 LTR: retaining functional coupling, of Nef and RBF-2, at RBEIII?* Int J Biol Sci, 2007. **3**(5): p. 318-27.
98. Estable, M.C., et al., *Distinct clustering of HIV type 1 sequences derived from injection versus noninjection drug users in Vancouver, Canada.* AIDS Res Hum Retroviruses, 1998. **14**(10): p. 917-9.
99. Koken, S.E., et al., *Functional Analysis of the ACTGCTGA Sequence Motif in the Human Immunodeficiency Virus Type-1 Long Terminal Repeat Promoter.* J Biomed Sci, 1994. **1**(2): p. 83-92.
100. Koken, S.E., et al., *Natural variants of the HIV-1 long terminal repeat: analysis of promoters with duplicated DNA regulatory motifs.* Virology, 1992. **191**(2): p. 968-72.
101. Chen, J., et al., *TFII-I regulates induction of chromosomally integrated human immunodeficiency virus type 1 long terminal repeat in cooperation with USF.* J Virol, 2005. **79**(7): p. 4396-406.
102. Reynolds, L., et al., *Repression of the HIV-1 5' LTR promoter and inhibition of HIV-1 replication by using engineered zinc-finger transcription factors.* Proc Natl Acad Sci U S A, 2003. **100**(4): p. 1615-20.