Introduction

The initiation of DNA replication

DNA replication is a vital cellular process controlled by cell cycle regulators which ensure accurate and complete duplication of the genome prior to cytokinesis. Replication is initiated by the activation of proteins loaded upon origins of DNA replication dispersed throughout the genome. Cell cycle regulators ensure that the activation of a given origin is limited to once per S phase, thereby preventing over replication and genomic instabilty. The initiation of DNA replication proteins are highly conserved and found in all eukaryotes studied to date: plants, yeast, flies, worms, fish, flogs, mice and humans [1]. Although regulation of initiation proteins is species specific, a model for the initiation of DNA replication has emerged (figure 1).

In 1964 Jacob et al., proposed the replicon hypothesis whereby binding of "initiator" proteins to genetic elements called "replicators" initiates DNA synthesis [2]. Experiments with purified *trans*-acting initiator proteins bound to *cis*-acting replicators or origins of DNA replication in bacteria, bacteriophages and eukaryotic viral systems resulted in a general mechanism for the initiation of DNA synthesis. In viral systems, a single origin is prepared for DNA synthesis usually through a single protein. For example, the SV40 large T antigen binds the viral origin, melts and unwinds the double stranded DNA allowing access of DNA polymerase α[3]. In *E. coli*, initiation is carried out with three proteins: DnaA binds the single bacterial origin, DnaC melts the double stranded DNA and DnaB promotes helicase loading [4]. The presence of a single genetically definable origin in both viral and bacterial systems allowed for the rapid elucidation of the proteins involved during the initiation process and provided the framework for experiments in yeast that have served as the basis of eukaryotic initiation (figure 1). The increase in genomic complexity of eukaryotes necessitates an increase in the number origins and the control of origin usage has been fine tuned throughout evolution to support DNA synthesis only once per cell cycle. The

importance of this is played out in metazoans where a balance between proliferation and differentiation is regulated by the genetic program of the cell through cell cycle regulators that drive proliferation, cellular division and differentiation.

In *Sacchromyces cerevisiae* origins of DNA replication (autonomously replicating sequences ARS), are *cis*-acting transposable sequences approximately 150 base pairs in length which when inserted into a plasmid can support extra-chromosomal maintenance [5]. To date, yeast ARS are the only genetically defined eukaryotic origins of replication and were used to identify the MCM proteins in yeast mutants [6]. In contrast, genetic metazoan origins remain elusive and are rather diffuse locations that can span thousands of base pairs where the replication complex assembles [7, 8].

In metazoans, the initiation of DNA replication begins at potentially thousands of origins upon which replication proteins are loaded forming what is known as a pre-replicative complex (pre-RC) [8, 9]. Pre-RC formation is achieved through the ordered loading of at least the following proteins: Origin recognition complex (ORC1 through 6), CDC6, CDT1, Minichromosomal maintenance complex (MCM2 through 7), MCM10 and CDC45. Cell cycle regulated activation of the pre-RC results in synthesis of DNA and traversal through S phase. During the early stages of S phase, origins are prevented from reinitiating synthesis through inactivation of various pre-RC components, such as CDC6, CDT1 and MCMs, by phosphorylation and subsequent translocation from the chromatin to the nucleoplasm.

The formation and activation of the pre-RC upon origins is best characterized through the model of "replication licensing" [10]. Replication licensing is the process by which potential origins are prepared for DNA synthesis through the ordered loading of proteins during G1 phase. Licensing begins early in G1 phase through the loading of ORC proteins onto origins. CDC6 and CDT1 are then

loaded and required for MCM complex loading. An origin is considered licensed when the MCM complex has been loaded. Cell cycle regulators ensure that once an origin is licensed and fired, DNA synthesis cannot be driven from the same origin until after the ensuing cytokinesis. Each of the various components is cell cycle regulated and thus present several layers control over origin firing. The proteins involved in pre-RC formation are introduced below.

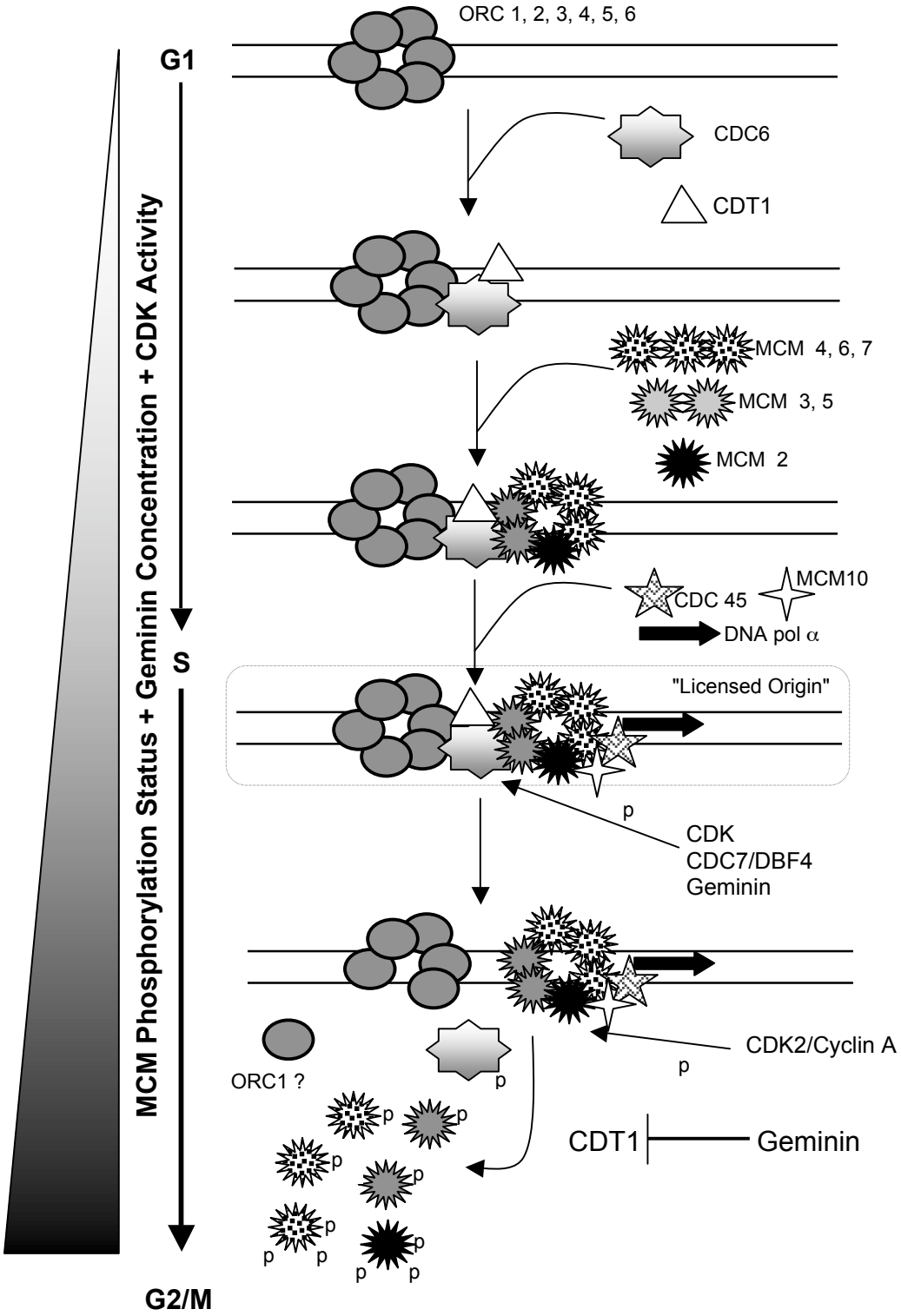

Figure 1. Model of the formation of the eukaryotic initiation complex. see next page for a detailed description.

Legend figure 1. Formation of the eukaryotic initiation complex. Late in mitosis and early in G1 phase potential origins of replication are marked by the assembly of remaining ORC subunits 2 through 6. CDC6 and CDT1 then bind through protein:protein interactions with ORC and through protein:DNA interactions. Soluble CDC6 is destroyed at the end of S phase while chromatin bound CDC6 remains . CDT1 binding along with CDC6 are necessary for MCM loading. MCM10 provides a bridge between the DNA and CDC45-DNA polymerase α. Once MCM proteins have been loaded, the origin is considered to be "licensed" for replication. CDC45, which facilitates the loading of DNA polymerase α, is then loaded and DNA synthesis begins. Once the origin has fired, cell cycle kinases inactivate the components of the pre-RC preventing relicensing of an origin until the next G1 phase. As S phase proceeds, MCM subunits are increasingly phosphorylated by CDKs resulting in their disassociation from chromatin. A number of MCMs remain bound thoroughout S phase perhaps required for late firing origins and/or to facilitate elongation. Geminin levels increase steadily throughout S phase, inactivating CDT1 and displacing it from chromatin. In the absence of chromatin bound CDT1, MCMs are unable to load onto potential origins. In G2 phase, the majority of MCM proteins are hyperphosphoylated and not bound to chromatin. At the end of mitosis, the pre-RC components begins to reassembly on potential origins.

Components of the Pre-Replicative Complex

The Origin Recognition Complex

The origin recognition complex (ORC) is composed of six subunits was identified by DNAse foot printing experiments on fractionated *S. cerevisiae* nuclear extracts [11]. ORC is conserved throughout all eukaryotic species [12]. The most well understood function of ORC is the binding to DNA which biochemically defines a potential origin. Although it is unclear that ORC plays a direct role in initiation of DNA replication, it is required for the loading of subsequent initiation proteins. ORC binds and hydrolyzes ATP [13]. While hydrolysis is not required for binding to DNA, it is necessary for downstream events in initiation such as DNA unwinding and interaction with CDC6. In *S. cerevisiae*, *S. pombe* and *D. melanogaster* ORCs are found constitutively bound to DNA throughout the cell cycle [14-17]. In *Xenopus* cells and egg extracts, ORC is displaced during metaphase and reforms upon origins early in G1 phase [16, 18]. In mammalian cells, ORC1 starts to accumulate in mid-G1 phase, and reaches a peak at the G1/S boundary [19]. ORC1 decreases to a basal level in S phase. In contrast, the levels of ORC subunits (ORCs 2-5) remain constant throughout the cell cycle [19, 20]. ORC1 degradation is achieved through the 26S proteosome which is blocked through protease inhibitors.

The majority of ORC mutations exhibit defects in DNA replication. Mutations in yeast orc2 and orc5, like the mcm mutants, do not support extrachromosomal maintenance [21]. ORC6 inactivation in HeLa cells and ORC2 inactivation in flies by RNAi results in drastic reductions in DNA synthesis and defects in chromatin condensation during M phase [22]. ORC2 hypomorphs generated in the human HCT cancer cell line by homolgous recombination exhibit an extended G1 phase, but proceed normally through S phase [23]. ORC1 inactivation in HeLa cells by RNAi decreases DNA synthesis and solubilizes chromatin bound ORC2 subunits [20].

CDC6

CDC6 (CDC18 in *S. pombe*) was identified in a screen for *S. cerevisiae* cell division cycle mutants and is a member of the large AAA+ family of ATPases [24]. Binding of ATP causes a conformational change in CDC6, which is thought to facilitate DNA binding . Its primary sequence is related to ORC1 and contains several phosphorylation sites in the N-terminus [25]. During G1 phase, CDC6 associates with CDT1 and ORC upon DNA and promotes MCM binding [26]. Immunodepletion of CDC6 in the *Xenopus* egg extracts prevents association of the MCM proteins and consequently the initiation of DNA replication [27]. In yeast, as cells pass into S-phase, CDC6 is phosphorylation by cyclin-dependent kinases (CDK), dissociated from the replication complex and degraded by ubiquitin-dependent proteolysis [28].

In mammalian cells, cdc6 expression is associated with proliferation and expressed during all stages of the cell cycle with a several fold increase in mRNA abundance at the onset of S phase [29]. Cdc6 expression, like most replication genes, is driven by E2F signaling and shut down in the absence of mitogenic stimulation [30]. Levels of CDC6 protein remain high throughout the cell cycle and its activity regulated by CDK mediated phosphorylation [31]. Microinjection of unphosphorylatable human CDC6 interferes with replication suggesting that it is required throughout S phase, its exact function during S phase is however unclear. Once inactivated CDC6 is transported out of the nucleus where it is ubiquinated and degraded by the APC [32].

CDT1 and Geminin

CDT1 was identified in *S. pombe* and functions with CDC6 in promoting the binding of the MCM complex [33]. Homologes have been identified in flies, frogs, humans and *S. cerevisiae*. Mutations in CDT1 result in a block to DNA replication and defects in S-phase check point [33, 34]. In *S. pombe* CDT1

expression peaks as cells enter S-phase, as does its inhibitor Geminin. As cells traverse S phase, CDT1 is inhibited by Geminin, which prevents its association with CDC6 and MCM proteins [35]. This action prevents loading of MCM proteins onto origins of DNA replication. During mitosis, Geminin is degraded via APC (anaphase promoting complex) mediated ubiquitination and subsequent proteosome 26 degradation thus allowing CDT1 to interact with CDC6 in the reloading of MCM proteins [36]. Recently, CDT1 has been shown to complex with mammalian MCM2 [37]. In quiescent, serum starved cells, association of CDT1 with MCM2 was low and upon restimulation increases, paralleling the loading of MCMs onto chromatin.

MCM

The MCM proteins were identified in yeast through genetic screens for mutations in plasmid maintenance and are proposed to be a DNA helicase when formed into a heterohexameric complex upon DNA [6, 38] . The MCM proteins are conserved members of the large AAA^{+} family of ATPases. Since their initial discovery in yeast, homologes have been described in all eukaryotes [1]. Studies of yeast null mutants have shown that each MCM subunit is absolutely essential for cell viability. The MCM holocomplex exhibits a robust ATPase activity and is required for MCM function. In yeast, ATP activity is mostly a function of the MCM3-MCM7 subcomplex and that MCM7 is largely responsible for ATP binding [39].

The licensing model for the initiation of DNA replication presents a model whereby origins are capable of supporting DNA synthesis only once per cell cycle [10]. Licensing is achieved by the cell cycle regulated loading of the MCM complex upon origins only after ORC, CDC6 and CDT1 are bound. After an origin has fired, the MCM complex is removed by phosphorylation and the origin inactivated. Studies in the *Xenopus* egg extracts have biochemically identified the MCM2-7 complex as the critical component in the licensing of the pre-RC

[40]. The proposed function of MCM complexes is based on *in vitro* observations of a helicase activity of human MCM4-6-7 subcomplexes that is abolished upon addition of MCM2 [41, 42]. Degron mutants in yeast at the restrictive temperature stall replication forks in S phase which suggests a function of MCM proteins not only during initiation but also for elongation [43].

In *S. cerevisiae* MCM proteins are nuclear in G1 phase and are translocated to the cytoplasm as S phase proceeds, a process that is mediated by CDK phosphorylation [44]. In metazoans, MCM proteins are constitutively nuclear [45-50]. During G1 phase the majority of the MCM proteins is associated with chromatin and hypophosphorylated. As S phase proceeds, MCM proteins become hyperphosphorylated and dissociate from the chromosome [42, 48, 51, 52]. MCM protein are found in abundance and their levels remain constant throughout the cell cycle while mcm gene expression peaks during the G1-S phase transition.

MCM10

MCM10 is an essential replication protein and was first identified in *S. cerevisiae* [53]. Homologes have been described in fission yeast, worms, flies, frogs and human. MCM10 is a nuclear protein and is localized to replication origins where it mediates the interaction between the MCM2-7 holocomplex with DNA [54-56]. RNAi inactivation of MCM10 in *D. melanogaster* resulted in chromatin condensation defect, chromosome loss and a reduction in DNA synthesis [57]. Additionally, MCM10 was shown to interact with ORC2, CDT1, MCM2 and CDC45. In human cells, mcm10 gene expression peaks during the G1-S phase transition and is found bound to chromatin throughout S phase. Mcm10 mutants show a drastic reduction in origin firing and an increase in stalled replication forks. Human MCM10 dissociates from chromatin in G2 phase after DNA synthesis has completed [58].

CDC45

CDC45 was originally identified in *S. cerevisiae* in a screen for cold-sensitive mutations in the cell cycle progression [59]. CDC45 is crucial for the transition of the pre-RC into a migrating replication fork. In yeast it has been shown to interact with MCM5 and MCM7 and appears to move with the replication fork [15, 56, 60-62]. In *Xenopus* egg extracts depletion of CDC45 prevents loading of DNA polymerase α [60]. Human CDC45 binds to MCM7 and the p70 subunit of DNA polymerase α, the only polymerase capable of *de novo* DNA synthesis [63]. In mammals, homozygous knockout mice are lethal [64]. As with the MCM proteins, CDC45 is phosporhylated by CDK and dissociates from chromatin at the end of S phase [60, 65].

Eukaryotic MCM proteins

The family of six conserved and homologous proteins, MCM2 through 7 is required for the initiation and completion of DNA replication. Three additional MCM proteins, MCM1, MCM8 and MCM10 have also been identified. MCM1 is a transcription factor that regulates the expression of mcm and other replication genes and is required for initiation [66]. MCM8 is a nuclear protein with an unknown function that interacts with the MCM complex [67, 68]. MCM10 interacts with the MCM2 through 7 complex stabilizing its interaction with DNA and with DNA pol α (see above). Mutations in all of the MCM proteins genes results in the loss of extrachromosomal plasmids carrying yeast ARS sequences [6]. Two-dimensional DNA gel analysis with MCM mutants where initiation events can be detected as intermediates that migrate as a ‘bubble” structure showed decreased initiation of DNA replication events. *Xenopus* egg extracts experiments wherein MCM proteins are immunodepleted prevent initiation of DNA replication[18, 40, 69].

Mcm genes have been cloned in *Schizosacharomyces pombe, Drosophila melanogaster, Xenopus laevis, Mus musculus, Homo sapiens*, plants and worms [1]. The sequence conservation of MCM proteins between yeast and humans is between 450 and 600 amino acids and thus are considered to be conserved throughout evolution. MCM proteins are abundant in proliferating cells and all subunits interact with each other. Indeed, immunoprecipitation of *Xenopus* egg extracts with antibodies to a specific MCM precipitates all the MCM proteins [70]. In human cells MCM proteins range in observed molecular weight from 85 to 125 kilo Daltons (kDa) and are constitutively nuclear [50, 71, 72].

In *Methanobacterium thermoautotrophicum* a single mcm gene exists and is most closely related to yeast and human MCM4 [73]. A putative zinc finger motive that is conserved in the MCM proteins found in Archeae and eukaryotic MCM2 suggests that MCM function evolved before the emergence of eukaryotes [74-76] .

In mammalian cells, mcm gene expression peaks during late G1 phase and is driven by E2F . The E2F transcription factor drives expression of replication genes involved in G1 to S-Phase transition [77, 78]. While mcm gene expression is periodic, MCM protein levels are constant throughout the cell cycle [79]. MCM proteins are found in abundance and have a half-life of approximately 21 hours [46, 80]. In differentiated and quiescent cells, mcm gene expression is turned off [48, 80, 81]. Thus MCM protein expression is a direct indicator of cellular proliferation and is increasingly used in the diagnosis of cancer, cardiac hypertrophy and other diseases that involve deregulation of proliferative control [82-87].

In yeast, the MCM proteins are more abundant than ORC proteins, transcribed during mitosis and early G1 and despite periodic transcription, MCM protein levels remain constant throughout the cell cycle [44]. Thus regulation of MCM function occurs at the level of binding to the pre-RC. MCMs are loaded onto

origins as hypophosphorylated species and as S phase proceeds, they are increasingly phosphorylated. In G2 phase MCMs are hyperphosphorylated and found in the nucleoplasm . Phosphorylation of MCM proteins is achieved through CDC7/DBF4 during S phase [88]. In *S. cerevisiae*, mutations in CDC7/DBF4 prevent S-Phase firing of late origins, whilst in *Xenopus* CDC45 association with the pre-RC is destroyed [88, 89].

Structural studies on the MCM2-7 complex in *S. pombe* have shown it to be a globular-ring shaped structure with a molecular mass of about 560 kD and possible a central cavity [90]. A complex containing human MCM4-6-7 was purified from HeLa cell extracts with a mass of 600kDa and exhibits a helicase activity [41]. Binding of MCM2 or MCM3/MCM5, destroys the helicase activity. To date a MCM complex composed of all six members, which exhibits helicase activity has not been purified.

All of the MCM subunits are vital as null mutants are lethal. That they are required for DNA replication has been described above. The initial identification of the MCM through yeast mutants described a general defect in S phase entry and decreased DNA synthesis [6, 24, 91, 92]. ORC, CDC6 and CDT1 have shown that they are no longer needed to support DNA synthesis after MCM loading [93-95]. In contrast, MCM degron mutants when shifted to the restrictive temperature exhibit stalled replication forks and a shut down of DNA synthesis [43]. Thus implicating a role of MCM proteins during elongation. RNAi studies in *D. melanogaster* have additionally shown defects in chromatin condensation in the absence of MCM2 CDC45 and MCM10 [57]. The chromatin condensation defect is also observed in MCM4, ORC2, ORC4 and ORC6 [22, 57]. Finally, MCM7 depletion in human HCT cancer cells by RNAi prevents DNA synthesis and S phase entry [96].

Goal of this thesis

The work undertaken in this thesis explores the role of MCM proteins during DNA synthesis in primary human cells. To date, there exists no *in vitro* assay to address the initiation of DNA replication in mammals. The majority of research on the initiation of mammalian DNA replication has been performed in HeLa cells wherein key cell cycle regulators are mutated or absent and S phase is uncoupled from mitogenic stimultion. Thus, I decided to carry out experiments in primary cells in the hopes of gaining insights into the initiation process as occurs in healthy human cells. Mouse knockout models are a valuable tool in dissecting gene function. However, to study the initiation proteins in such models would be a laborious and highly risky project. DNA replication genes are needed at the earliest stage of embryogenesis and classical knockout mice in the homozygous state would be lethal at the earliest stages of embryogenesis. Indeed, this was observed for the mouse cdc45 gene [64]. As an alternative, I decided to use phosphorothioated antisense oligonucleotides which have been shown to effectively shut down gene expression in many cell types including primary human cells [96-98]. Experiments will be centered on the targeted down regulation of the mcm4 gene.

The mcm genes 2 through 7 encode proteins that form a complex upon origins of DNA replication having a proposed DNA helicase function [1]. The MCM4 protein is a member of a subcomplex (MCM4-6-7) containing an *in vitro* helicase activity [41]. MCM4 is a target for MCM complex regulation by the CDKs [88]. MCM4 may also represent the archetypical MCM protein [74]. MCM mutants exhibit a variety of cell cycle effects, most of which prevent S phase entry. I will address three main questions utilizing a combination of antisense technology, immunoblotting of fractionated cell extracts, microscopy and flow cytometry. Experiments will be carried out in a primary human fibroblast cell line, IMR90.

Questions to be addressed

1. What is the effect of blocking mcm4 gene expression on the remaining mcm gene products at the mRNA and protein level?

I will look at the effect of antisense mediated down regulation on mcm4 and several other mcm gene products at the mRNA level by Real-time rtPCR. In addition, I will explore the effects at the protein level by immunoblotting.

2. How is the MCM complex affected in the absence of a specific sub-unit?

MCM proteins form a heterohexameric complex upon chromatin. I will determine the effects of targeted MCM4 down regulation on the formation of the MCM complex by fractionation of cellular proteins and immunoblotting.

3. What are the cell cycle effects in response to targeted down regulation of mcm gene expression?

MCM proteins are essential for DNA synthesis and traversal through S phase.
The effects on DNA synthesis and cell cycle distribution in response to targeted MCM4 down regulation will be determined by flow cytometric methods and flourescent microscopy.