ISOLATION AND GENETIC DISSECTION OF A EUKARYOTIC REPLICON THAT SUPPORTS AUTONOMOUS DNA REPLICATION

A Thesis

by

SHIBANI DATTA

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

December 2005

Major Subject: Genetics

ISOLATION AND GENETIC DISSECTION OF A EUKARYOTIC REPLICON THAT SUPPORTS AUTONOMOUS DNA REPLICATION

A Thesis

by

SHIBANI DATTA

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee,	Geoffrey M. Kapler
Committee Members,	Linda Guarino
	David O. Peterson
Chair of Genetics Faculty,	James R. Wild

December 2005

Major Subject: Genetics

ABSTRACT

Isolation and Genetic Dissection of a Eukaryotic Replicon That Supports Autonomous DNA Replication. (December 2005)

Shibani Datta, B.S., University of Mumbai;

M.S., University of Mumbai

Chair of Advisory Committee: Dr. Geoffrey M. Kapler

Maintenance of genome integrity requires that chromosomes be accurately and faithfully replicated. We are using *Tetrahymena thermophila* as a model system for studying the initiation and regulation of eukaryotic DNA replication. This organism contains a diploid micronucleus and polyploid macronucleus. During macronuclear development, the five diploid chromosomes of the micronucleus are fragmented into 280 macronuclear minichromosomes that are subsequently replicated to ~45 copies. In stark contrast, the 21 kb ribosomal DNA minichromosome (rDNA) is amplified from 2 to 10,000 copies in the same nucleus. Previous characterization of the rDNA replicon has led to the localization of its origin and the cis-acting regulatory determinants to the 1.9 kb 5'non-transcribed spacer region.

The objective of this study was to identify and characterize non-rDNA origins of replication in *Tetrahymena*. This will help determine the underlying basis for differential regulation of rDNA and non-rDNA origins during development, as well as provide a better understanding of the organization of eukaryotic replicons. To this effect, I

developed a DNA transformation assay that I used to isolate new *Tetrahymena* replication origins. A 6.7 kb non-rDNA fragment, designated TtARS1, was shown to support stable autonomous replication of circular plasmids in *Tetrahymena*. Genetic dissection revealed that TtARS1 contains two independent replicons, TtARS1-A and TtARS1-B. Full TtARS1-A function requires a minimal sequence of 700 bp, and two small regions in this fragment have been shown to be essential for origin function. TtARS1-B replicon function was localized to a 1.2 kb intergenic segment that contains little sequence similarity to TtARS1-A. Both non-rDNA replicons lack sequence similarity to the rDNA 5' NTS, suggesting that each replicon interact with a different set of regulatory proteins. This study indicates that the rDNA and the non-rDNA replicons have a modular organization, containing discrete, cis-acting replication determinants.

DEDICATION

This thesis is dedicated to my loving parents, and my brother Debashis, for their undying love, patience, and support throughout this process. They have always strived to put me on the right path and without them I would not be here today. You are a part of every page, every thought.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Geoffrey Kapler, for his enthusiastic supervision, guidance and support throughout the duration of this project. He has encouraged me to take on a challenging project and helped me work through the stumbling blocks along the way. I sincerely thank you for the invaluable training I have received while in your laboratory. I also thank you for always having your door open for questions and discussions.

I would like to thank my committee members, Dr. Linda Guarino and Dr. David Peterson. I thank them for their time and effort they have put in to improve this thesis. I also acknowledge you for the stimulating discussions that we had during committee meetings. I feel that each of you has contributed significantly, and uniquely, to my continuing development as a scientist and I thank you for the knowledge that you have given me.

I would like to especially thank my colleagues and the graduate students in the lab, Sebastian Yakisich for scientific discussions, Tara Morrison for her guidance and help with cloning, Mohammad Mohammad for discussions and all the teasing and fun in the lab, Taraka Donti for all the support, Pamela Sandoval for being a great friend, and most of all, for all the help with experiments, slides and everything else, Nandina Paria for being a great co-worker, and Aaron Smith for all the help. You are an incredible group with whom I had several productive discussions, debates and laughter and you are the reason that it will be hard for me to leave Texas A&M University.

I would like to thank my friends in India and in the United States for the moral support, encouragement and their faith in me. A special thanks to Dr. Dorothy Shippen and all the members of Shippen lab for the stimulating discussions at lab meetings, and the lab instruments that I have borrowed from you.

Finally, I would like to thank those closest to me, my family, whose presence helped me throughout my career. I owe a debt of gratitude to my parents and my brother Debashis that I won't be able to express. No matter what happened, you have always been there for me, cheering me all the way and sharing my ups and downs. It's been a long road, and I've been a long way from home, but with your help I have been able to achieve my successes. Your constant inspirations, sacrifices and hard work have meant the world to me. I love you all, and thank you.

Finally, I would like to thank Surajit Dasgupta, for being a great friend. You have always motivated me to succeed in my pursuits, have had faith in me and provided me with all the support at the good times and the bad times. You have always believed in my ability to succeed even when I thought it wasn’t possible. I appreciate and value all that you have done for me. Thank you for all that and more. To all others that I have not mentioned, you are greatly appreciated.

TABLE OF CONTENTS

CHAPTER Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

DNA replication is a fundamental and highly orchestrated process in all living organisms. In order to maintain the genetic integrity of organisms it is required that the genome be accurately and faithfully replicated. In eukaryotes, DNA replicates during the S-phase of the cell cycle. Replication initiates at specific sites in chromosomes called origins of replication. Replicons are comprised of specific cis-acting determinants to which trans-acting factors bind. They basically consist of three components, a binding site for the origin recognition complex (ORC), an easily unwound sequence called as the DNA unwinding element and binding sites for non-ORC regulatory proteins, including transcription factors (reviewed in 1). Multiprotein complexes are assembled and two replication forks are formed at each origin for bidirectional replication (2). Figure 1.1a shows bi-directional replication, starting from a central initiation site or origin. Although the basic regulation of DNA replication has been elucidated, the organization of replicons differs from species to species. While a number of eukaryotic replicons have been identified, the exact cis-acting determinants for origin function have been characterized in only a handful of them. While the critical trans-acting factor, ORC, is conserved in eukaryotes, the cis-acting DNA sequence requirements for ORC targeting and regulation are not.

This thesis follows the style of *Nucleic Acids Research.*

The replicon model of DNA replication

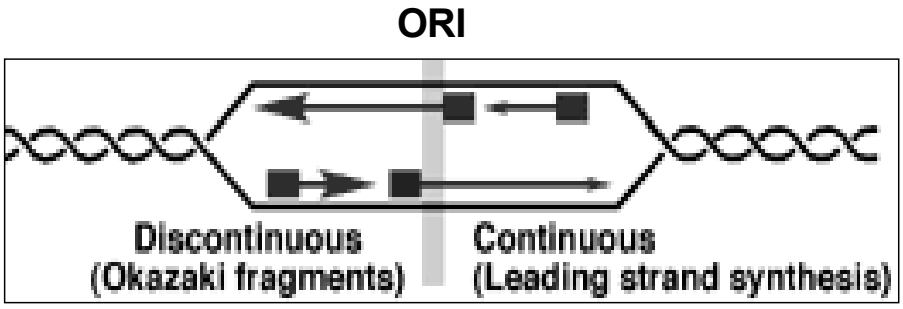
The replicon model was first proposed by Jacob, Brenner and Cuzin (3). It is based on the idea that specific DNA target sequences, called the replicators, serve as binding sites for trans-acting regulatory factors called initiators. Once the initiator is bound, other replication-associated proteins, such as DNA polymerase and primase, are recruited and the replisome is formed. The replicon model has been used to explain the process of replication initiation in many organisms, including viruses and bacteria.

In order to test whether the replicon model fits eukaryotic DNA replication, experiments were done with the budding yeast, *Saccharomyces cerevisiae*. Replicators were identified by their ability to mediate the extrachromosomal replication of plasmids. Initiators were subsequently discovered on the basis of their ability to bind to replicators. The most important initiator protein that has been discovered is the six subunit origin recognition complex (ORC) (4). ORC is conserved in all eukaryotes, from yeast to man (5), suggesting that the initiation mechanism should be essentially the same in all eukaryotes. ORC acts as the landing pad for the pre-replication proteins.

Organization of eukaryotic replicons

Eukaryotic genomes have a higher complexity than prokaryotes, both in terms of size and organization. Prokaryotes as a rule have a single circular chromosome, in contrast to the eukaryotes which contain multiple linear chromosomes. Prokaryotic chromosomes typically contain a single origin of replication. In contrast, eukaryotic origins are generally spaced 50-100 kb apart (6), with tens to hundreds of origins in a

1.1a



1.1b

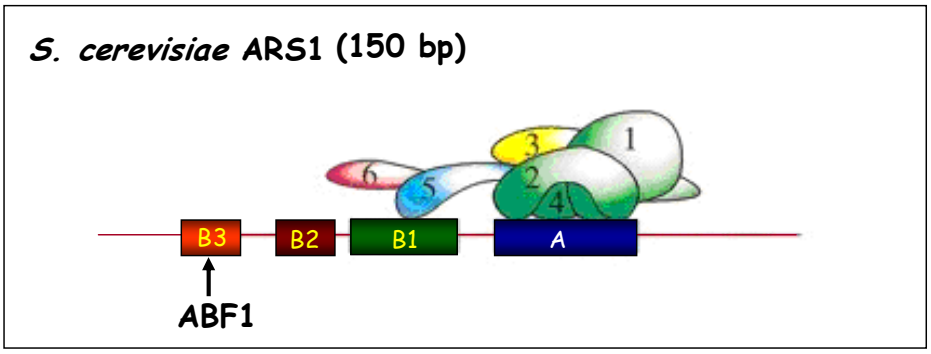


Figure 1.1 DNA replication in eukaryotes.

Fig.1.1a represents duplex DNA undergoing bidirectional replication from a centrally placed origin (ORI). Fig. 1b is a schematic representation of S. cerevisiae ARS1 replicon. ARS, B1, B2 and B3 are cis-acting sequences in the replicon. Red ovals represent nucleosomes. ORC or origin recognition complex binds to ARS sequence and B1 element. B2 is a DUE (DNA unwinding element) and B3 is the binding site for transcription factors like Abf1. The entire replicon is contained within 150 bp of nucleosome free region in the DNA.

single chromosome. Eukaryotic replicons are strategically present in AT-rich regions of the genome, such as intergenic or non-coding segments. Several trans-acting proteins and local chromatin structure affect the initiation of replication. It is due to this complexity that the study of eukaryotic replication initiation has been challenging. The identification of origins through loss-of-function assays in native chromosomes is difficult, since the inactivation of a single replicon is compensated for by the firing of neighboring origins. Most of our current understanding of eukaryotic replication origins is based on the use of organisms that are easy to manipulate genetically and can harbor artificial minichromosomes.

Saccharomyces cerevisiae *replication origins*

The pioneering model in which eukaryotic replication origins were first identified is the budding yeast, *Saccharomyces cerevisiae*. Replicons were identified as DNA sequences that conferred autonomous DNA replication to plasmids. These autonomous replicating sequences (ARSs) are distributed in the yeast genome at a frequency that is consistent with the spacing of origins estimated from electron microscopy (7). Minichromosome maintenance assays allowed the identification of many yeast origins. Once these origins were identified, two dimensional gel electrophoresis was used to study the firing of these origins in their native chromosomal locus.

Genetic dissection of one such autonomous replicating sequence, ARS1 (8), revealed a modular array of short functional cis-acting elements that co-localized with the initiation site. A graphical representation of *S. cerevisiae* ARS1 is in Figure 1.1b. As

shown, the entire ARS1 replicon spans ~150 bp and contains multiple discrete cis-acting determinants, one of which is a conserved 11-bp segment, the ARS consensus sequence present in all functional origins. This 11 bp ACS (ARS consensus sequence) is essential for binding of the origin recognition complex (ORC) (9).

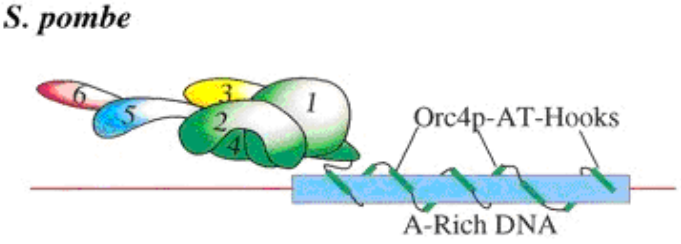
ARS1 consists of the essential A element within which is the 11 bp ARS sequence, and three additional elements, designated as B1, B2 and B3 (Figure 1.1b). The A element by itself is not sufficient for origin function. Each B element by themselves is not essential; however, together they are necessary for origin function (8, 9, 10, 11). The A and B1 elements of ARS1 replicon constitute an approximately 30 bp segment that forms the bipartite binding site for ORC. The B2 element forms the DUE that also possesses weak ORC binding activity. Element B3 is an auxiliary component that binds to a bi-functional protein, ABF1, which serves as a transcription factor at other sites in the genome and a replication protein at the ARS1 locus. B3 can be replaced by binding sites for other transcription factors (8). Thus, the ARS1 replicon has a modular structure with distinct cis-acting determinants that regulate replication initiation. Systematic dissection of another replicon in *S. cerevisiae*, ARS121, has shown the presence of specific cis-acting functional domains. In *S. cerevisiae* it has been observed that about 15-20% of the sequence is shared amongst all origins, which basically comprise of A and B1. Non-ARS regulatory determinants are generally: interchangeable between origins (12).

Schizosaccharomyces pombe *replication origins*

The fission yeast *Schizosaccharomyces pombe* represents a distinctly different paradigm for eukaryotic replicons. Like *S. cerevisiae*, origins were identified in *S. pombe* by their ability to confer autonomous replication property to plasmids (13). However, in contrast to the small replicons of *S. cerevisiae*, *S. pombe* replicons are larger, about 500 to a few thousand basepairs in length. Some *S. pombe* replicons contain a match to the 11 bp *S. cerevisiae* ACS, however, this sequence is not essential for origin function. In fact, no consensus sequence has been identified. Instead, multiple redundant sequences are present, which can be interchanged with different AT-rich sequences without significantly affecting origin function (14, 15). *S. pombe* replicons have a very high A+T content, which is often distributed asymmetrically with A residues on one strand and T residues on the other strand. Like most origins, *S. pombe* replicons are mostly present in intergenic regions. However, a difference is that they are preferentially located in regions containing promoters of divergent transcription units (16).

Since *S. pombe* replicons lack any consensus sequence for recognition, there has to be some way to determine origin specificity. One of the means is through the association of SpORC with the origin. A graphical depiction of *S. pombe* origin recognition is shown in Figure 1.2a. Studies have revealed that SpOrc4 protein contains nine copies of a sequence motif termed the AT-hook (17, 18). This segment binds to AT-rich stretches of DNA (Figure 1.2a). *In vitro* studies have shown that the interaction of SpOrc4p with DNA does not require either ATP or any other ORC subunit (18). This

1.2a



1.2b

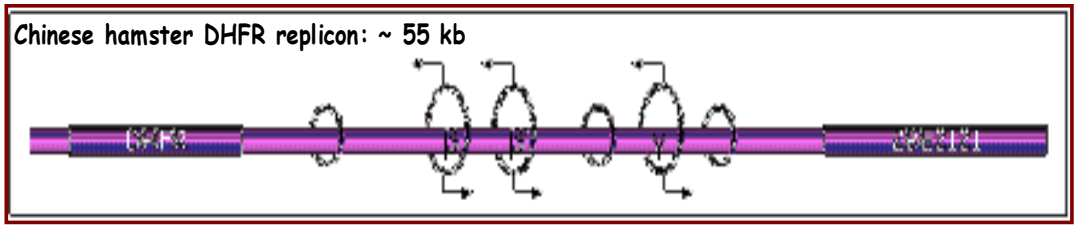


Figure 1.2 Organization of replicons in eukaryotes.

Fig. 1.2a is a description of origin specification in *Schizosaccharomyces pombe,* where SpORC4 determines origin specificity. SpORC4 subunit interacts with degenerate AT-rich sequences of DNA and then the rest of the ORC complex binds to the SpORC4. Figure 1.2b shows the Chinese hamster dihydrofolate reductase replication origin. There are multiple sites of initiation within a broad intergenic region of 55 kb, between the DHFR gene and 2BE2121 gene of unknown function.

probably negates the necessity for a consensus sequence for replicon function. Moreover, this AT hook is not present in Orc4 orthologs in other model eukaryotes. Furthermore, unlike *S. cerevisiae*, there is no known role for non-ORC DNA binding proteins in origin specification. Thus, the fission yeast follows a more stochastic model of ORC association and origin firing than the classical replicon model, indicating that there are different modes for origin recognition by ORC (19).

Metazoan DNA replication origins

The identification of metazoan replicons has been a slow process compared to yeast. Part of the reason is the high complexity in metazoa where replication initiation is determined by several parameters in addition to DNA sequence. Moreover, as compared to yeast, metazoan replicons are much larger in size. The genetic dissection of metazoan replicons has been especially challenging, where just one or two replicons have been dissected per model species. These replicons contain discrete cis-acting genetic determinants, but there are few common features between them. Furthermore, the different replicons show no specificity for ORC binding *in vitro*. Also, there are genetic elements that interact with non-ORC binding proteins. Some of the metazoan origins that have been characterized are described below.

Chinese hamster dihydrofolate reductase replication origin

A graphical representation of the Chinese hamster DHFR replicon is shown in Figure 1.2b. Replication of the DHFR locus occurs within a very broad initiation zone of about 55 kb. Origin activity starts from three major sites: ori β, ori β' and ori γ (20, 21, 22, 23). Ori β is approximately 17 kb downstream from the DHFR gene, closely followed

by ori β', while ori γ is more distally placed at 23 kb downstream of the gene. Ori β region is capable of initiating DNA replication in ectopic chromosomal locations. Neither ori β' nor ori γ are necessary for initiation at ectopic loci (24). However, in the endogenous chromosome, two dimensional gel electrophoresis showed that most initiation events occur within a 12.5 kb region that contains both ori β and ori β' (25). Also, in the endogenous locus, a 3.2 kb fragment present at the 3' end of DHFR gene is required for all initiation activity in the 55 kb DNA region (24). This can be attributed to insulation from position effects that can prevent initiation to occur.

Like other eukaryotic replicons, DHFR ori β is rich in AT sequences. In addition, ori β contains a stably bent DNA sequence (26) and a deletion of 4 bp in a GC hexanucleotide near this sequence reduces replication initiation (24). Whether this effect is due to altered spacing between the cis-acting determinants or due to the reduction in size of a GC stretch is not known. These studies indicate a modular organization of the DHFR replicon with an AT-rich DUE and a requirement for specific sequences either to maintain nucleosome spacing or for the binding of replication initiation proteins.

Human β globin replicator

The expression of the human β globin gene is regulated in a tissue-specific manner, where replication occurs during early S phase in the erythroid cell line and in middle-to-late S phase in non-erythroid cell. This makes the β globin locus an attractive model for studying the determinants of replication initiation sites, as well as replication timing. A pictorial representation of the human β globin is shown in Figure 1.3. The endogenous human β globin origin spans a DNA segment of over 60 kb (27, 28),

consisting of an initiation region (IR) and a locus control region (LCR) that is located approximately 50 kb away from the IR. The IR can function as a replicator at ectopic loci and hence contains the genetic determinants for the initiation of replication. The LCR is not required for ectopic initiation (29). However, in the endogenous locus the LCR is necessary for origin activity. Studies have revealed that LCR plays a role in determining the timing of replication of human β globin in different tissues (30). It has been hypothesized that in the native chromosome, the LCR probably renders an open configuration to chromatin, which is necessary to promote early origin firing (29, 30).

Human lamin replication origin

A replication origin has been discovered within a 1.2 kb segment of DNA in human chromosome 19 which overlaps the 3' non-coding end of the lamin B2 gene and the promoter of the adjacent gene (31). This segment can function as a replicator at ectopic loci. This is by far the shortest reported human sequence that can act as a replicator. Extensive mapping of this replicon has shown that there are two crucial components, a 290 bp sequence that shows protection in DNA footprint analysis and nearby G+C rich island (CpG island) (32, 33). Extensive mapping has localized the origin of bidirectional replication to 3 bp sequence within the protected region (34).

Human c-myc replication origin

The replication origin at the human c-myc locus has been localized to a 2.4 kb region which is 5' to the c-myc gene. Replication starts at this location and several other sites over a 12 kb segment encompassing the c-myc gene in the endogenous chromosome (35, 36). The 2.4 kb fragment contains a DNA unwinding

1.3

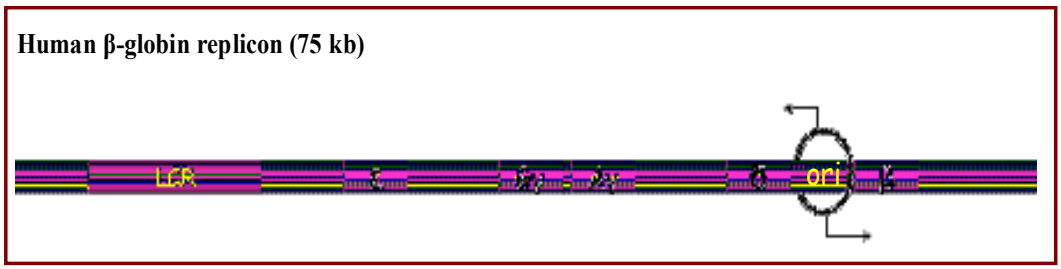


Figure 1.3 Organization of the human β-globin replicon.

The human globin replicon spans a DNA segment >75 kb. Replication initiates within a few kilobases between the adult δ and β-globin genes. LCR is locus control region, located >50 kb away from the initiation region. LCR is necessary for all initiation events from the globin locus in the endogenous chromosome.

element (37). This 2.4-kb c-myc origin fragment confers autonomously replicating sequence (ARS) activity on plasmids, which specifically initiate replication in the origin fragment *in vitro* and *in vivo* (38). Later, studies showed that this segment can also function as a replicator at ectopic sites in the chromosome (39). Therefore, this fragment contains all the cis-acting determinants necessary for replication initiation from human c-myc origin. Thus, from the above examples, it is evident that mammalian replicons appear to consist of discrete cis-acting determinants, like those in the case of *S. cerevisiae* origins. The complexity of mammalian replicons, however, is much higher and sequences that are a long distance away from the initiation region can have profound influence on origin function. In spite of the presence of defined elements that direct replication initiation, biochemical and plasmid maintenance studies have suggested that mammalian ORC does not exhibit sequence specificity. Plasmid transformation revealed that any piece of DNA, irrespective of the source, can confer autonomous replication property to plasmids (40). This probably suggests that no consensus sequence is necessary for replication initiation in mammalian systems.

Drosophila chorion gene locus

Drosophila melanogaster is one of the most genetically tractable metazoans wherein replicons can be studied. The genes encoding for the chorion (eggshell) proteins in *Drosophila* are present on two chromosomes, the X chromosome and chromosome 3. To meet the high demand for chorion proteins during oogenesis, the follicle cells amplify the two chromosomal domains containing the chorion gene clusters to ~16 and ~60 fold respectively. This amplification occurs due to repeated firing of replication origins in the

chorion gene cluster (41). The origin of replication in chromosome 3 chorion locus has been studied in greatest detail. A pictorial representation of *Drosophila* chorion gene replicon is shown in figure 1.4. This replicon spans an approximately 10 kb segment of DNA containing cis-acting determinants, which include essential elements as well as auxiliary elements. An essential determinant is the α-element, which is upstream of the S18 chorion gene; another essential element is the β-region which is downstream of the S18 gene. Both have multiple repeats, high AT content and similarities to the yeast consensus sequence (42, 43, 44). Genetic dissection of these two regions revealed that most of the 320 bp α-element is essential for amplification and it has been called the amplification control element (ACE3) (45) (Figure 1.4). Approximately, 70-80% of the initiation events occur near the β-region and hence the region was named ori β (46). Mutational analysis narrowed the critical elements to a 142 bp sequence within ACE3 and a 140 bp segment plus a 226 bp AT-rich sequence in ori β (44). In addition to the ACE3 and ori β, there are other cis-acting determinants that promote gene amplification, which are not absolutely essential. These include two less frequently used origins, the ori α and ori γ and four amplification enhancing regions (AER-A, AER-B, AER-C and AER-D) (47) (Figure 1.4).

Experiments where ACE3 and ori β have been inserted in exogenous chromosomal locations have shown that these elements are sufficient to drive amplification (48, 49). Additionally, *in vivo* studies indicated that *Drosophila* ORC binds to ACE3 and ori β. However, quantitative *in vitro* DNA binding studies of DmORC revealed that it binds to origin (ACE3 and ori β) and non-origin sequences with similar

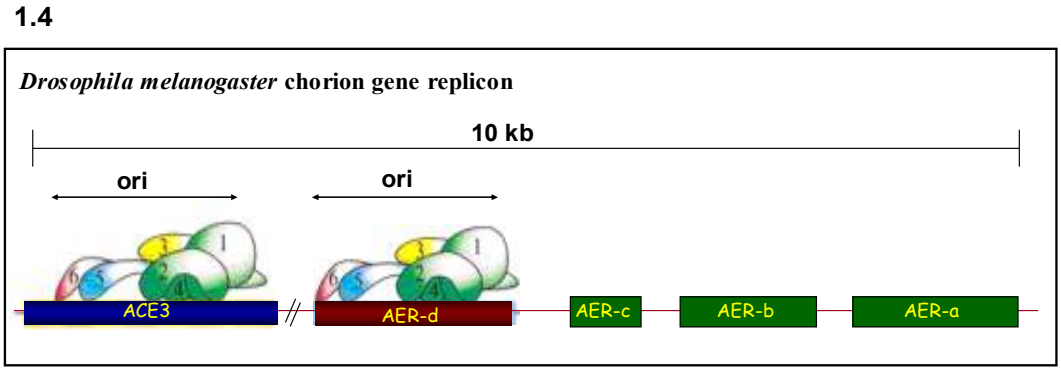


Figure 1.4 Organization of the *Drosophila* chorion gene replicon.

The chorion gene replicon spans a DNA segment of 10 kb, and consists of cis-acting determinants such as amplification control element (ACE3) and amplification enhancer elements (AERa-d). AER-d is also called ori β and it serves as the initiation site or origin. ACE3 and ori β are binding sites for ORC and are essential for replication.

affinities (50). This suggests that DmORC does not bind in a sequence-specific manner and hence, most plasmids transfected into *Drosophila* tissue culture cells undergo autonomous replication (51).

rDNA replicon in Tetrahymena

The above example described a developmentally regulated replicon in *Drosophila melanogaster*. Another replication origin that undergoes locus specific amplification is the one present in the ribosomal DNA minichromosome of *Tetrahymena thermophila*. The rRNA genes of *Tetrahymena thermophila* reside in a small 21 kb macronuclear chromosome that has two head-to-head copies of the ribosomal RNA gene and flanking DNA sequences (52, 53). Thus one copy of the entire rDNA is 10.5 kb in length. Figure 1.5 depicts the differential DNA replication during macronuclear development of *Tetrahymena.* The RNA genes, which are present in a single copy in the germline micronucleus, reside in a large chromosome. They are flanked by chromosome breakage sequences (Cbs) (54) (Figure 1.5). The micronuclear rDNA copy is removed from the chromosome during macronuclear development by site specific fragmentation at the Cbs elements. Palindrome formation occurs by a mechanism that requires short inverted repeat at one end and involves intra-molecular recombination (55, 56). Following palindrome formation, locus specific amplification occurs, wherein the rDNA copy number in the developing macronucleus increases to 10,000 copies (Figure 1.5).

Figure 1.6 depicts the rDNA replicon in *Tetrahymena*. The rDNA replicon resides in the 1.9 kb 5' non-transcribed spacer region (5' NTS). This segment contains

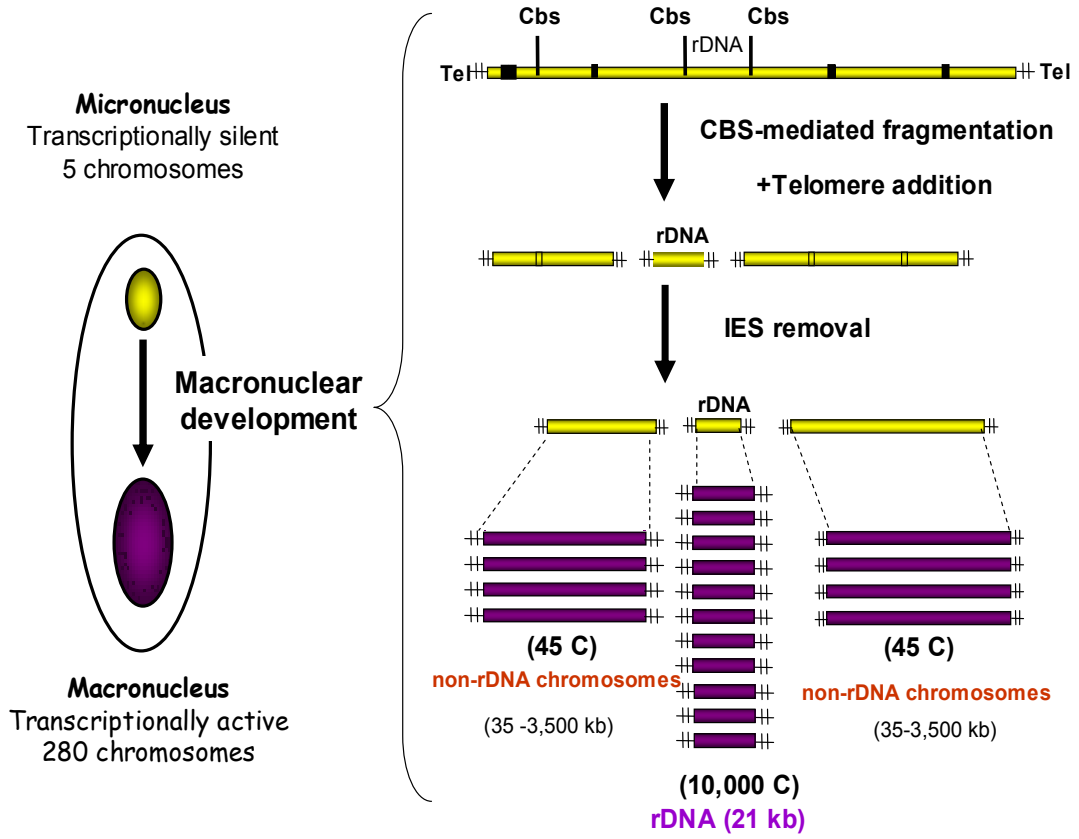


Figure 1.5 Differential DNA replication during macronuclear development of *Tetrahymena*.

The left panel is a pictorial representation of bi-nucleate *Tetrahymena* cell containing the germline micronucleus and a transcriptionally active macronucleus. Macronuclear development is a key step in *Tetrahymena* sexual cycle. As shown in the right panel, during macronuclear development, the micronuclear chromosomes are fragmented at chromosome breakage sites (Cbs), telomeres are added at the ends and internally eliminated sequences (IES) are removed by chromosome breakage and fusion. Endoreplication of the genome occurs to generate 45 copies of the non-rDNA chromosomes in the macronucleus. Locus specific amplification of the rDNA occurs at this point and the rDNA copy number in the developing macronucleus increases to 10,000 copies.

all the cis-acting determinants necessary for amplification as well as vegetative replication of linear and circular rDNA episomes (57, 58, 59). The initiation sites localize within two distinct 430 bp domains within the 5' NTS called domain 1 and domain 2. These initiation sites are part of a single replicon. This means that even though there are two closely spaced origins, only one of them fires in a given molecule. Furthermore, mutations in domain 2 affect replication initiation from both origins. Domian 1 and domain 2 are not sufficient for replication initiation by themselves, indicating that adjacent cis-acting elements are also necessary (60).

Molecular and genetic approaches have been used to identify cis-acting determinants for rDNA replication. Phylogenitically conserved elements include the type I elements, type II elements, type III elements and the pause site elements (Figure 1.6) (59, 60, 61, 62). Another cis-acting determinant of rDNA replication is the replication fork barrier, but it is not conserved. Of the phylogenetically conserved sequences, the type I elements and the pause site elements have been shown to control replication initiation. The type II and the type III elements are also present in the 5' NTS (61), but their effect on replication is not known. There are four type I elements in each 5' NTS, one in each domain (type IA & IB) and two at the promoter (type IC & ID). Minichromosome maintenance studies suggested that mutations within or downstream of the type I elements cause defects in initiation of replication (59, 62, 63). These findings indicate that replication initiation is regulated by long distance DNA–protein / protein–protein interactions (60).

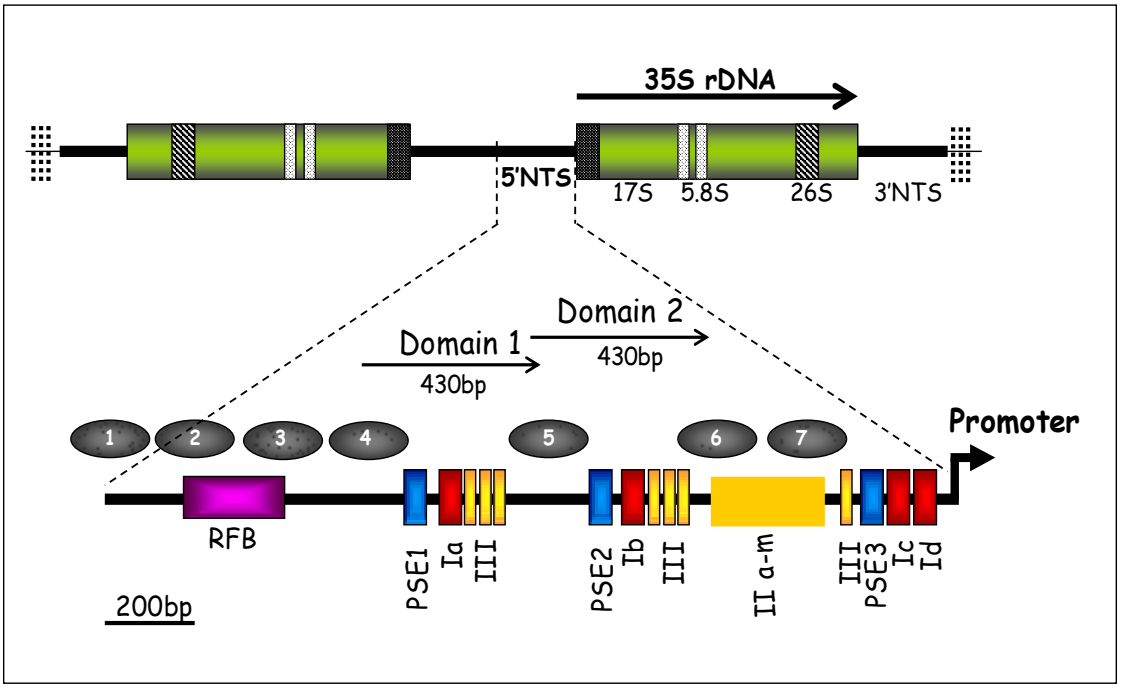


Figure 1.6 Structural and functional features of the *Tetrahymena* rDNA minichromosome.

The rDNA replicon resides within 1.9 kb 5' non-transcribing spacer region (5'NTS). It consists of cis-acting determinants like Type Ia-d elements, type III elements, the pause site elements (PSE) and the replication fork barrier (RFB). There are two origins, one in domain 1 and the other in domain 2. The origins are localized within 430 bp of nucleosome free regions.

The second genetically defined cis-acting replication determinant present in the 5' NTS is the pause site elements (PSE) (Figure 1.6). These are 52 bp segments of DNA where replication forks arrest transiently (64). There are three PSE, which are present in the nucleosome-free regions and shared blocks of sequence similarity. Transformation studies where the 52-bp tripartite PSE3 element was substituted with a DNA fragment of identical length and comparable A+T content indicated that the PSEs are essential for maintenance of the plasmids (65). This suggested that PSE3 contains an intact cis-acting element that is necessary for replication. Another cis-acting functional determinant affecting rDNA replication is the replication fork barrier (RFB) (Figure 1.6). These are sites where replication forks arrest irreversibly (66). Experiments using two dimensional gel electrophoresis have suggested that the RFB is active in amplifying DNA molecules and it blocks the progression of replication forks near the center of the rDNA palindrome. RFB differs from pausing sites in that they block any further movement of a replication fork rather than just transiently arrest its progression.

In addition to the cis-acting determinants in the rDNA replicon, there are trans-acting replication proteins that regulate replication initiation. Four different type I element binding factors, TIF 1 through 4 have been identified in *Tetrahymena*. *In vitro* experiments have revealed that these factors preferentially bind to single strand DNA (67, 68, 69). TIF1 is a homotetramer that contacts DNA via a single 24 kDa subunit (68). It binds to the A-rich strand of the type I element at the rDNA origin and the T-rich strand of the type I element at the rDNA promoter. It has been suggested that TIF1 might stabilize the replication origins in their unwound state (68). TIF1 also associates with the

pause site elements, both *in vitro* and *in vivo* (65). Both these interactions are necessary for initiation of replication, thereby suggesting that TIF1 selectively modulates the binding of a different subset of trans-acting factors to the respective regulatory cis-acting elements (65). Recent data has revealed that *Tetrahymena* mutants that are deficient in TIF1 exhibit early firing of rDNA replication origin, an elongated macronuclear S-phase and delayed cytokinesis (70). The observation that TIF1 depletion accelerates the timing of rDNA origin activation suggests that TIF1 is a negative regulator of rDNA replication

TIF2 and TIF3 have been identified on the basis of their ability to bind to type I element sequences *in vitro* (71). TIF2 has a molecular mass of 85 kDa and its binding activity increases in cells undergoing vegetative replication or rDNA amplification. In sharp contrast to this, TIF3 levels are highly elevated in non-replicating cells and its DNA binding activity is lost when rDNA replication begins. This suggests that it may be a negative regulator of rDNA replication. It has a molecular mass of 32 kDa (71).

TIF4 is a unique ORC-like complex which binds in an ATP-dependent sequence specific manner to the T-rich strand of the type I elements (69). It is present in both the micronucleus and the macronucleus. TIF4 exhibits chromatin association, origin-specific DNA binding, and is organized in a multi-protein complex that contains at least one putative ORC subunit (antisera to *Xenopus* Orc2 recognize a 69 kDA subunit in the TIF4 complex). However, the mode of DNA recognition by TIF4 is different from *S. cerevisiae* ORC, since it binds to single strand DNA. The DNA binding activity of TIF4 is regulated in a cell cycle-dependent manner, peaking in the S-phase. Thus the type I elements interact with an ORC like (TIF4) and non-ORC (TIF1-3) complex, and this

dynamic interplay may suggest a novel mechanism for regulation of rDNA replication initiation (69). Recent work has revealed that TIF4 is a ribinucleoprotein (RNP). It has a RNA subunit, designated as 26T, which has been mapped to 280 nucleotides in the 3' terminal of 26 rRNA (72). Chromatin immunoprecipitation experiments with tagged 26T RNA indicated that the RNA localizes to the rDNA origin. Also, the *in vitro* DNA binding of TIF4 to DNA is dependent on the RNA subunit, and this is signifies a novel mechanism for the targeting of proteins to DNA.

The rDNA replication origin is the only characterized replicon in *Tetrahymena* and it has a modular structure similar to *S. cerevisiae* replicons. It is developmentally regulated and undergoes locus specific gene amplification. In order to elucidate the mechanism of amplification and the cis-acting determinants necessary for this event, comparative analysis is required with other replication origins in *Tetrahymena* that are not amplified. To serve this purpose, in my thesis research I set out to identify new *Tetrahymena* replication origins. I adopted the method of identifying sequences that confer onto plasmids the ability to support autonomous replication in *Tetrahymena*. Once such a sequence was identified, the next step would be to genetically dissect the non-rDNA replicon to find cis-acting elements necessary for origin activity. While a number of replication origins have been identified in eukaryotes, few have been characterized genetically. *Tetrahymena thermophila* is a very amenable model eukaryote for this study since it can propagate both natural and artificial rDNA minichromosomes as linear or circular episomes, the later of which could be used to facilitate the isolation and characterization of new replication origins.

Regulation of origin activation

In the above section, I have described eukaryotic replicons and have provided examples of the most well-characterized replicons in different species, with emphasis on the cis-acting functional determinants. However, a number of trans-acting factors are also necessary for the initiation of DNA replication, which takes place in distinct steps. These steps include the formation of pre-replicative complex, the pre-initiation complex and the post-replicative complex.

Overview

When eukaryotic cells proliferate, origins of DNA replication alternate between a competent and non-competent state. An overview of the licensing of DNA replication is shown in Figure 1.7. DNA replication initiation in eukaryotic cells begins with the formation of the pre-replicative complex (pre-RC) during late M phase and continues in G1 phase of the mitotic cell cycle. The first set of initiator proteins assemble on the origin in a stepwise fashion, and these include ORC, Cdc6, Cdt1 and Mcm2-7. The second step is the maturation of pre-RC to form the pre-initiation complex (pre-IC), when the cyclin-dependent kinases are active and another set of initiator proteins are recruited to the origin. This event is marked by the association of Cdc45 protein with the origin and recruitment of DNA polymerases. After initiation of DNA replication from an origin, the pre-IC is dismantled and the origin switches to a non-competent state, called the post-replicative state (post-RC). The events and the proteins associated with the formation of pre-RC, pre-IC and post-RC are described in more detail below.

Pre-replicative complex

The formation of pre-replicative complex (pre-RC) is the first step in eukaryotic DNA replication (Figure 1.7). This process is also called origin licensing. The various components involved in the formation of pre-RC are as follows:

Origin recognition complex: At the heart of the pre-RC lies the most important protein, the six-subunit origin recognition complex (ORC), which acts as a landing pad for other replication proteins like Cdc6, Cdt1, and the MCM2-7 complex (73, 74, 75). ORC is a conserved feature of all eukaryotes, where it acts as the initiator at origins of replication. ORC interacts with cis-acting determinants at the origin and can determine origin specificity. In *S. cerevisiae*, ORC binds to double strand DNA, interacting primarily with the A-rich DNA strand of the A and B1 elements of ARS1 replicon *in vitro* (4, 76) binding of which requires ATP. Single strand DNA binding leads to the activation of ATPase activity, which in turn causes a conformational change in ORC (77). Once bound, ORC is associated with origins in *S. cerevisiae* throughout the cell cycle (78). In contrast to this situation, ORC binds to *S. pombe* replicons in a stochastic manner. The SpOrc4 subunit interacts with AT-rich DNA in a non-specific mechanism by means of its AT-hook, and this binding specifies origin activity (17). Mammalian and *Drosophila* ORC subunits lack the AT-hook, but bind DNA with no apparent specificity *in vitro*, similar to SpORC. Thus the mechanism of origin specification and recognition is different in different organisms, and the complexity of this process cannot be explained by the simple replicon model.

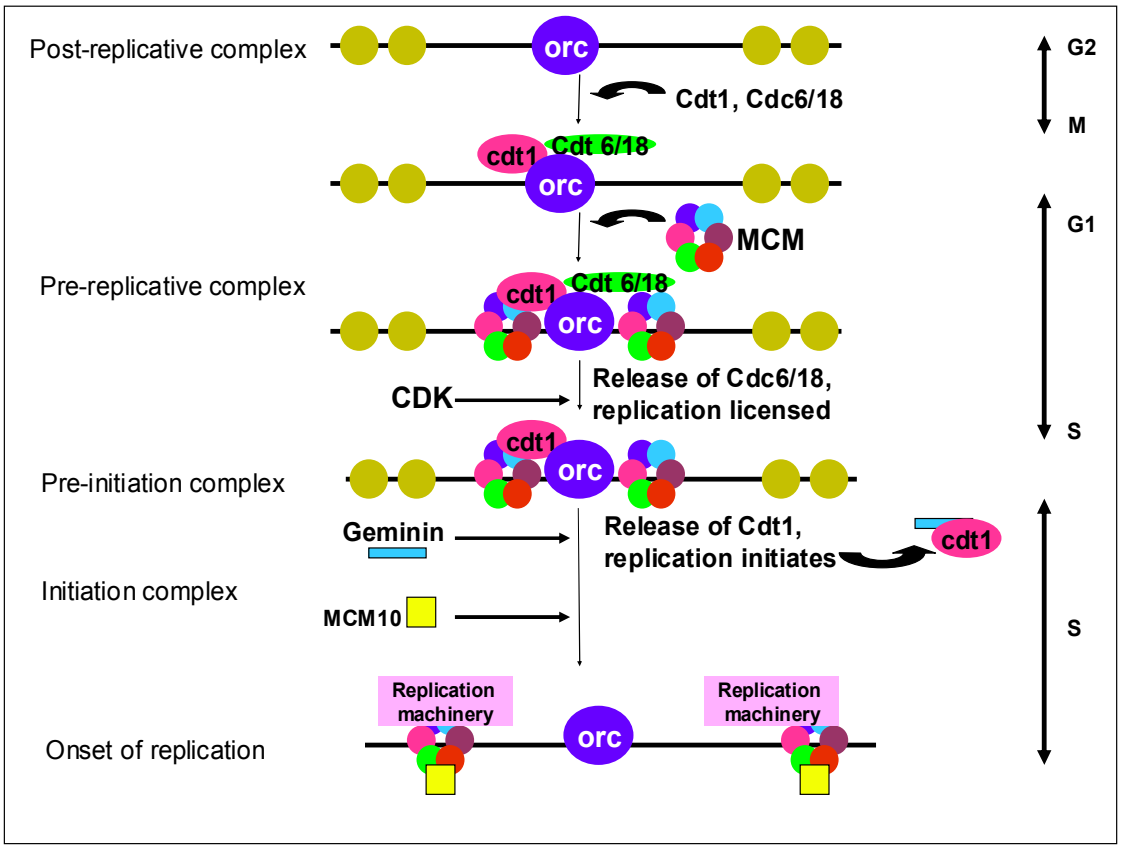


Figure 1.7 Licensing of eukaryotic DNA replication involves the step-wise assembly of initiator proteins on origins of replication.

During the G1/S phase, assembly of pre-replicative complex occurs. ORC acts as a landing pad for Cdc6/18 and Cdt1, which load the MCM complex on to the chromatin. This is followed by the dissociation of Cdc6/18, which is degraded by CDK activity. Geminin, present from the onset of S phase to the end of M phase, binds to Cdt1 and prevents licensing. MCM10 interacts with the rest of the MCM complex at the origins and this marks the formation of initiation complex.

Other components of pre-RC

Cdc6 or cell division cycle 6 protein: Cdc6 was first discovered in a screen for temperature sensitive cell division control mutants that exhibited a defect in the initiation of DNA replication (79). In budding yeast, Cdc6 binds to chromatin near origins in late M to early G1 phase through an interaction with ORC or another origin-associated factor (80). Cdc6 in association with Cdt1 plays a role in recruiting the MCM complex onto origins and thus regulates the timing of the pre-RC formation.

Cdt1: This was first discovered in *S. pombe,* and has been found in all examined eukaryotes. Cdt1 is necessary not only for the nuclear accumulation of the Mcm2-7 proteins, but also for their loading onto chromatin (81). It has a very important regulatory role to ensure that replications initiates once and only once at each origin per cell cycle. It was observed that overexpression of Cdt1 in *Xenopus* was sufficient in causing re-replication at origins (82).

Minichromosome maintenance (MCM) proteins: The MCM protein complex consists of six heterologous proteins, Mcm2-7. They were discovered in a study involving *S. cerevisiae* mutants that were defective in the replication of minichromosomes that contained a single replicon (83). Studies involving the purification of replication licensing factor complex from *Xenopus* eggs revealed that the MCM2-7 complex is its active component (84). Mcm4/6/7 form a double-trimeric complex which possesses DNA helicase activity (85, 86). This complex is necessary for the unwinding of DNA *in vitro* and thus plays an important part in the progression of replication fork (87). Once MCM proteins are loaded onto the origins, the origins are

licensed for replication. Recently, a new protein, Mcm10, was discovered for its role in initiation of DNA replication in *Saccharomyces cerevisiae*. The mcm10-temperature sensitive mutant shows defects in maintenance of minichromosomes and induces cell cycle arrest of cells with 2C DNA content. Mcm10 physically interacts with proteins from MCM2-7 family (88). *In vivo* studies have shown that Mcm10 is an essential part of the replication fork, migrating with the fork after origin firing, as well as regulating the stability of Cdc17, the catalytic subunit of the DNA polymerase α-primase complex (89). Interestingly, recent *in vitro* data in *Xenopus* indicates that MCM10 binds to origins of replication after the MCM2-7 complex and recruits Cdc45 to the origin (90). A regulatory protein MCM8 has been recently discovered, which functions during replication elongation and not initiation. MCM8 associates with chromatin only after the initiation of replication. *In vitro* studies revealed that MCM8 has DNA helicase activity and DNA–dependent ATPase activity (91, 92).

Cyclin dependent kinases & Cdc7/Dbf4 complex: After the loading of MCM complex on the origin of replication, the stage is set for the action of cyclin dependent kinases (CDK) and Cdc7 family kinases. One of the protein complexes that play a role in regulating the initiation of eukaryotic DNA replication is Cdc7/Dbf4 complex. It is a protein kinase which is required for initiation to occur at each replication origin (i.e. early in S phase for the initiation of early-firing origins as well as late in S phase for the initiation of late-firing origins) (93, 94). Cdc7/Dbf4 kinase is essential for the recruitment of Cdc45 on the origin. *In vitro* studies have indicated that MCM2-7 complex, Orc4, DNA polymerase α and other replication proteins are *in vitro* substrates

for phosphorylation by the Cdc7/Dbf4 complex (81, 95). However, the *in vivo* substrates of Cdc7/Dbf4 protein kinase are not known. Moreover, this complex can associate with the origin prior to MCM recruitment as in the case of *S. cerevisiae*, or after the association of MCM as is the case in Xenopus egg extracts (96, 97).

Pre-initiation complex

The pre-replicative complex matures to form the pre-initiation complex during S phase of the cell cycle. Once the MCM complex is bound at the origin, Cdc45 associates with ORC and sequences flanking the origin. A new complex is formed at this point, referred to as the pre-initiation complex. Cdc45 protein is required for the association of DNA polymerase with chromatin and is required for both replication initiation and elongation (98, 99,100). Once the pre-IC is formed, the origin starts to unwind followed by loading of replication protein A. Subsequently, DNA polymerase is recruited and replication initiates (101, 102).

Post-replicative complex

Once DNA replication has initiated from an origin, the pre-IC is dismantled. Certain initiator proteins such as ORC or ORC subunits may remain bound to the origin, however, the pre-RC cannot be re-assembled until mitosis and cell division has occured. The re-assembly is prevented by the action of CDKs. This non-competent state of an origin of replication is called post-replicative complex or post-RC. The post-RC is formed during the S/G2 phase of the cell cycle.

Thus, it is evident and not very surprising that the process of DNA replication is highly regulated and is under strict cell cycle control. The various proteins that have

been mentioned above play an important role, but in addition to these, there are other regulatory mechanisms to modulate origin function. Some of these mechanisms are mentioned below.

Regulation of replication initiation

Eukaryotic DNA replication is regulated at various stages of replication initiation by specific mechanisms and a brief description of the different regulatory steps are given below.

Cdc6/Cdc18 complex

One mechanism of regulation of DNA replication is through the phosphorylation of Cdc6/Cdc18 by cyclin-dependent kinase activity. The kinase inhibits the assembly of the pre-initiation complex at the replication origin. In *S. cerevisiae*, phosphorylation of Cdc6 protein with cyclin-dependent kinase leads to activation of the replication origin and further causes degradation of Cdc6 (103). Phosphorylated Cdc6 also loses the ability to interact with the origin and pre-RC assembly and re-initiation is prevented. Yeast mutants that lack the motif for interaction with CDK fail to grow in the absence of wild type Cdc6 (104). Similarly, mutation of CDK phosphorylation sites in the *S. pombe* Cdc18 protein leads to increased Cdc18 activity and this in turn increases re-initiation of replication from origins *in vivo*. Thus CDK dependent phosphorylation inhibits Cdc18 function (105). The regulation of Cdc6/Cdc18 by CDK is conserved in humans, where phosphorylation of Cdc6 by CDK activates the origin, inhibits Cdc6 function and causes its export from the nucleus (106, 107).

MCM regulation

It has been suggested that CDK phosphorylation of MCM proteins can inhibit their association with chromatin. In *S. cerevisiae, in vivo* data indicate that reinitiation from origins is prevented by blocking pre-RC assembly in part by promoting nuclear export of MCM proteins (108, 109). Dislodging of MCM from the DNA and their subsequent phosphorylation by CDK is essential for this export. *In vivo* and *in vitro* studies have revealed that *Xenopus* MCM4, which is a part of a large protein complex comprising several other MCM proteins undergoes CDK dependent phosphorylation. This event reduces the affinity of MCM4 for chromatin and this may be a mechanism for inactivation of MCM complex (110, 111). In humans, MCM2 and MCM4 are substrates for CDK (cdc2 and cyclin B) *in vitro* and it has been suggested that phosphorylation of MCM complex by cdc2 kinase may be one of mechanisms that negatively regulates the MCM complex-chromatin association, thereby preventing re-initiation (112).

The metazoan ORC cycle

The origin recognition complex plays an important role in ensuring that the initiation of replication occurs once and only once per cell cycle from each origin. In *S. cerevisiae*, ORC remains bound to the origins of replication throughout the cell cycle, however, in S and M phase of cell cycle it gets phosphorylated by CDK and this inhibits its assembly into pre-RCs. In mammals, there is an additional layer of regulation to prevent origin re-firing via the dissociation of ORC subunits. In the G1 phase of cell cycle, the ORC holo-complex is bound to the origins. On entry into the S phase, when replication has initiated, the Orc1 subunit is selectively destabilized from the complex,

released from chromatin and subsequently degraded in an ubiquitin-dependent manner (113, 114). In a special case such as *Drosophila* oogenesis, when cells reach the 10B stage and genome wide endoreduplication stops and locus specific amplification of chorion gene starts. At this point, Orc1 and Orc2 localize to form discrete foci that are sites of gene amplification. Once replication ceases, Orc2 leaves the foci and cannot be detected (115, 116). *In vitro* studies with *Xenopus* egg extracts indicated that *Xl*Orc1 and *Xl*Orc2 are bound to chromatin throughout interphase but gets displaced from chromatin during mitosis, at least partly (117).

Epigenetic regulation of eukaryotic DNA replication

Relationship between chromatin structure and origin function

A handful of metazoan origins have identified, but it is still elusive as to how certain segments of DNA are selected as sites for the initiation of DNA replication. No consensus sequences have been found to predict origin activity and metazoan ORCs show no apparent sequence specificity *in vitro*. Recently, studies have shown that one of the factors that contribute to the temporal regulation of mammalian origins of replication is the higher order structure of chromatin (118, 119). It has been established that transcriptionally active euchromatic regions replicate during early S phase, whereas silent heterochromatic regions replicate in late S phase (120, 121, 122).

An example of the link between histone modification and chromatin remodeling with DNA replication is the Epstein Barr virus origin of plasmid replication (OriP). ORC is recruited to OriP via interactions with the DNA binding viral protein, EBNA (123). OriP is flanked by nucleosomes, which are subject to remodeling in late G1 phase via

histone actelylation. This event coincides with the recruitment of ATP-dependent chromatin-remodeling factor SNF2h to OriP and MCM loading. Thus cell cycle-dependent changes in chromatin structure affect the assembly of pre-RC at OriP (124).

Studies using P-element mediated transformation in *Drosophila* also revealed that the amplification of chorion genes is very sensitive to their position in the genome (44, 49, 125). Histones at the active chorion gene origins are hyperacetylated, and this is coincident with the binding of ORC. When a histone deacetylase was tethered to the chorion locus 3 origin, the activity of the origin decreased, although the effect was quite modest (126). A recent study has shown that the firing time of *S. cerevisiae* origins can be artificially advanced either by global hyperacetylation of the genome due to the deletion of a histone deacetylase or by local recruitment of a histone acetyltransferase, suggesting that histone acetylation is a direct determinant of the timing of origin activation (127).

CpG islands: CpG islands are short G+C-rich regions of DNA that can be subjected to methylation (128). The methylation state of CpG sites can affect origin firing and hence it serves as an epigenetic control of DNA replication. CpG islands are approximately 1 kb long and are found at the promoters of many mammalian genes (129). Recent studies have revealed that certain CpG islands are initiation sites for both transcription and DNA replication in Chinese hamster ovary cells (130, 131). Generally methylation of CpG islands leads to silencing of specific genes. *In vitro* studies with *Xenopus* have revealed that the presence of methylated CpG islands inhibits the binding of ORC and initiation of replication (132). In Chinese hamster cells at the DHFR locus, a

greater than 50% reduction in CpG methylation at ori-β caused significant loss of activation of the origin. Therefore, CpG DNA methylation does affect the initiation process at some origins (133).

Effect of transcription on DNA replication

It has been observed that transcriptionally active loci are replicated early in S-phase, and late replication is correlated with transcriptional inactivity. The cause and effect relationship between transcriptional inactivity and late replication has been studied in yeast (134), where it was revealed that juxtaposition of a transcriptional silencer, such as HMR-E, with an early firing origin delays the timing of initiation from that origin. This occurs in the presence of Sir proteins, which are silencers of transcription. An effect of transcription factors on origin function is observed in *Xenopus*, where site specific DNA replication can be induced by the assembly of a transcriptional domain (135). Deletion of the promoter abolished initiation of replication. The relocalization of the transcriptional domain to an ectopic site induced a new origin of replication. However, active transcription is not required for this to occur, just transcription complex assembly is necessary. An interesting hypothesis is that transcription complexes are a determinant for initiation site specificity, and it may be due to this reason that during early development in *Xenopus*, there is no specificity of origin firing (135). These data indicate that transcription factors play a role in modulating origin activity.

Methods for mapping and studying replication origins and replicons

The identification of origins of DNA replication in eukaryotes has been a very arduous task. Prokaryotes generally have a single, clearly defined segment of DNA that

contains all the necessary sequences for origin activity. Eukaryotes however, have multiple origins, ranging from hundreds to often thousands. Eukaryotic replicons may or may not contain obvious sequence motifs and they are very difficult to characterize genetically. Moreover, the firing of different origins is temporally regulated, with early and late firing origins. In many cases there are broad initiation zones that span several kilobases. The selection of different origins in organisms can be influenced by the kind of nutritional media the cells are growing on (136). Recently, there has been evidence of epigenetic control of origin firing. With all the mentioned factors that determine origin function, it is very challenging to identify eukaryotic origins of replication. However, with the onset of many new techniques that have become available, a number of eukaryotic origins have been identified, and some have been well characterized genetically. An origin of replication can be defined both functionally (physical location where DNA synthesis occurs) and genetically (cis-acting determinants for replication). Thus there are two methods to identify origins; the first is based on the physical mapping of origins by analysis of nascent strands and replication intermediate structures. The second method consists of functional assays for identifying cis-acting regulatory elements.

Methods based on nascent strand analysis

1) Earliest labeled DNA fragment: In this method, a synchronized population of cells is released in the S phase in the presence of radiolabelled DNA precursor such as ^{14}C labeled thymidine. This is followed by cell lysis and DNA extraction. The DNA is digested with restriction enzymes and separated on a gel. The earliest labeled strands

are identified by autoradiography as the newly replicated nascent DNA. This technique was used to study the initiation of DNA replication in the amplified Chinese hamster DHFR locus, where distinct series of restriction fragments could be observed in genomic digests separated on ethidium bromide-stained agarose gels (20). Autoradiography of Southern blots of labeled genomic digests showed that DNA synthesis initiated in a small subset of the EcoRI fragments derived from the amplified units. These early labeled fragments are not synthesized at later times during S phase. Alternatively, the radiolabelled nascent strands can be used to probe cloned genomic DNA sequences (137). The earliest labeled DNA fragments are the ones that are closest to the origins. This is the simplest method to locate origins. However, the cell population must be well synchronized, and the there should be several hundred copies per genome of the DNA region of interest.

2) **Replication origin trap:** This technique is based on crosslinking of DNA followed by *in vivo* replication in the presence of radioactive precursors. The nascent DNA at newly activated replication origins get trapped between the crosslinks and can be separated from the bulk genomic DNA by centrifugation through an alkaline sucrose gradient (138). The nascent strands can be cloned and sequenced, or they can be used as probe against cloned genomic DNA fragments. This method enables the isolation of a representative population of chromosomal origins. However, one source of interference comes from repeated DNA (eg. Alu repeats) that can produce non-specific signals. Also, crosslinking can alter cell physiology, such that replication initiates at sites that are not normally used.

3) **Nascent strand extrusion:** In this method a synchronized population of cells is released into the S phase in the presence of radioactive DNA precursors. The newly synthesized DNA gets radiolabelled. Following this, the nuclei are isolated and DNA synthesis is continued in the presence of Hg-dCTP. The nascent strands get radiolabelled and this property is used for identification. Moreover, the strands also are labeled with Hg-dCTP affinity label that is useful for purification by affinity chromatography. After extrusion, the complementary nascent DNA strands anneal to give double stranded DNA, which can be cloned and sequenced, and further confirmed for *in vivo* origin activity (139).

4) **Nascent strand abundance assay:** In a log phase culture, the elongating replication forks moving away from any given origin contain nascent strands. In this method, the nascent DNA strands are isolated and fractionated on the basis of their size on an alkaline sucrose gradient. The shortest of the nascent strands maps closest to the origin. The abundance of nascent strands can be measured by competitive PCR or real time PCR assays. The primers used for competitive PCR are specific for the DNA region of interest. This method is extremely sensitive as well as quantitative, and allows for the detection of single copy sequences. It has been used for fine structure mapping of the Chinese hamster DHFR origin (24) and human c-myc locus (39).

Method based on the analysis of the structures of replication intermediates

1) **Two dimensional gel electrophoresis:** This method relies on the aberrant migration of DNA fragments that contain replication bubbles (sites of bidirectional replication)

in agarose gels. It provides knowledge on the firing of replicons in the native chromosome. It has been used successfully on *S. cerevisiae*, *S. pombe, Tetrahymena thermophila* and also on higher eukaryotes such as *Drosophila melanogaster,* humans and Chinese hamster ovary cells. The DNA is digested with an appropriate restriction enzyme, and run on an agarose gel. The first dimension separates the DNA fragments on the basis of size. The DNA is then run on a second dimension agarose gel, where the separation is on the basis of shape and size. Replication intermediates migrate slower than linear duplex DNA due to their mass and shape. Using radiolabelled probes one can detect the replication intermediates. A bubble arc is indicative of sites of bidirectional replication from a centrally placed origin. If the DNA is passively replicated from a distant origin, then a Y-shaped arc is observed. Modification of gel composition can be done to determine the direction of the movement of replicating forks. This method can be used to map initiation sites at the level of single copy sequences.

Methods based on the functional analysis of origin activity

1) **Assays for autonomous replication**: This assay is also called as ARS assay. It is based on the principle of the presence of cis-acting determinants for origin function. Any DNA segment that contains these cis-acting functional elements would be able to confer autonomous replication property to plasmids. The eukaryote in which ARS assay was first exploited is the budding yeast, *S. cerevisiae* (140). Plasmids containing a selectable yeast marker were used as vectors to clone yeast genomic DNA segments, followed by transformation into yeast cells. ARS containing

plasmids produce high frequency transformation. This is a fast and reliable method for the identification of cis-acting sequences that function as origins in a limited number of species. However, in metazoans this technique is more challenging. Centromeres in *S. cerevisiae* are small, approximately 100 bp in size. Hence, it is easier to do minichromosome maintenance studies. However, in case of humans, the centromeres are much larger (about 40 kb). This larger size precludes the construction of minichromosomes. Therefore, in the ARS assay, sequences can be lost due to mis-segregation and this makes the application of ARS assay to metazoan system tedious. Also, sometimes integration of plasmids can contribute to errors. In the ARS assay, the DNA region of interest is studied in an extrachromosomal context and sometimes this change in chromatin environment may cause the firing of cryptic origins.

In mammalian cells, short term ARS assays are often combined with a restriction enzyme approach such as Dpn*I* digestion. Plasmids containing potential ARS sequences are generated in bacteria that methylate their DNA. The plasmid is then transformed into mammalian cells, most of which lack DNA methylation. Any ARS sequence that replicates in the mammalian cells will be hemimethylated and finally unmethylated if it undergoes more than one round of replication. These replicated DNA molecules will be resistant to Dpn*I* endonuclease (which only digests fully methylated DNA). Thus, whether or not, a sequence has origin function can be determined.

2) **Genetic assays for defining replicons in mammalian systems and higher eukaryotes:** Many eukaryotic replicons have been functionally analyzed by placing a series of candidate replicators at ectopic sites in the chromosome, however in some cases the chromosomal context of a replicator might influence its ability to initiate replication. Such position effects can be overcome by placing a series of replicator candidates at constant genomic sites. The human β-globin replicon was studied by placing the 8.0 kb replicator fragment at ectopic loci in the simian genome, and was further genetically dissected by deletion analysis (29). This method was also used for the functional analysis of the human c-myc replicon. It comprised of transferring the origin at ectopic loci and further mapping the origin by nascent strand abundance assay and competitive PCR (39). Another method to characterize replicons is by placing them between chromatin insulators. The cis-acting elements of the *Drosophila* chorion gene replicon were identified by P-element transformation assays. In this assay, the 10 kb chorion gene replicon was inserted at ectopic sites in the genome, in between large DNA fragments that served as buffers (43). The results indicated the presence of amplification enhancer regions that are localized within 2.3 kb DNA sequence. To further study sequence requirements for chorion gene amplification, a vector was used in which transgenic constructs of the replicon are protected from chromosomal position effects by flanking insulator elements. The results showed that ACE3 and ori β are necessary and sufficient for amplification, and that ACE3 activates ori β in cis (49). Another strategy that has been used to study replicator activity is the analysis of pools of transformed cells (24). To identify

the cis-acting determinants of the DHFR replicon, the 5.8 kb fragment containing ori β was introduced into random ectopic chromosomal locations in a hamster cell line that lacked the endogenous DHFR locus. Pools of transformants were analyzed by competitive PCR based nascent strand abundance assay, which showed that specific sequences in the 5.8 kb region are necessary for initiation of replication.

Tetrahymena as a model organism

Biology of Tetrahymena

In my thesis research, I employed *Tetrahymena thermophila* as a model eukaryote to study DNA replication. The genus *Tetrahymena* belongs to the group of ciliated protozoa, which are an early branch of unicellular eukaryotes. *Tetrahymena thermophila* is the most well characterized species in this group and it is a large, motile, phagotrophic organism, which normally inhabits streams, lakes and ponds. The normal dimensions of this organism are 50 μm in length and 20 μm in maximum width. It is holotrichous ciliate, which means that it possesses cilia all over its surface. The genome size is approximately 1 x 10^8 bp. It has a short doubling time of about 2.5 - 3 hours in rich media and can survive within a temperature range of 12 to 41^{o}C (141).

A stark feature of ciliates is nuclear dimorphism, which means the presence of more than one type of nucleus in a single cell. *Tetrahymena thermophila* contains two structurally and functionally distinct nuclei, the smaller germline micronucleus and the larger transcriptionally active macronucleus. The micronucleus is diploid with five pairs of chromosomes which are transcriptionally silent (141). It is the store for genetic information that is transmitted to the progeny. The micronucleus divides by conventional

mitosis and meiosis. The larger nucleus is the somatic macronucleus, which is expressed during vegetative growth. The macronucleus is polyploid in nature and contains about 280 chromosomes. The macronuclear chromosomes lack centromeres and divide amitotically. They are derived from the five pairs of micronuclear chromosomes (142, 143). This process of formation a new macronucleus from the smaller micronucleus is called macronuclear development. No macronuclear DNA is transmitted to sexual progeny.

Cell division in *Tetrahymena* occurs by binary fission during vegetative growth (144, 145). *Tetrahymena* also undergoes a sexual cycle through the process of conjugation. During conjugation, two cells pair and a temporary conjugation tube is formed. Exchange of pronuclei occurs through this tube. For conjugation to occur the cells must be starved of at least one essential nutrient, be of different mating type and be sexually mature (141). In nature, conjugation seems to be an evolutionary method to propagate progeny under conditions of deprivation. In the laboratory, conjugation of *Tetrahymena* can be induced by starving the cells for several hours in a sterile dilute salt solution, and then mixing cells of different mating type. Figure 1.8 is a detailed pictorial representation of nuclear events that occur during conjugation in *Tetrahymena*. The key stages in conjugation are stated below:

1) Pair formation
2) Two rounds of meiosis to generate four micronuclei
3) Selection of one micronucleus and degeneration of the other three micronuclei

4) Mitotic division of the selected micronucleus followed by reciprocal micronuclear exchange between the two cells in the conjugating pair
5) Pronuclear fusion in each mating partner to form progenitor micronucleus
6) Two rounds of post zygotic divisions occur to generate four micronuclei
7) Two of the four micronuclei develop into two macronuclei by macronuclear development (macronuclear anlagen). At this point each cell in the mating pair has two micronuclei and three macronuclei of which one is the old macronucleus
8) Pair separation, degradation of the old macronucleus and one micronucleus (at this point each cell has one micronucleus and two macronuclei)
9) Micronuclear mitosis and cell division to generate progeny with one micronucleus and one macronucleus

Macronuclear development in *Tetrahymena* involves site specific fragmentation of the five pairs of micronuclear chromosomes at chromosome breakage sites (Cbs). Following this event, telomeres are added to the ends of the newly generated fragments (146, 147). The Cbs sequences are thus present in the micronucleus only and are about 17 bp in length. The micronuclear DNA also contains additional sequences called internally eliminated sequences (IES) that are removed during the formation of macronuclear chromosomes by chromosome breakage and rejoining. The bulk of the macronuclear DNA is endoreplicated to represent 45 copies per macronucleus during macronuclear development (145). There is one exception however. During the process of macronuclear development, the micronuclear copy of the ribosomal DNA gets excised and rearranged

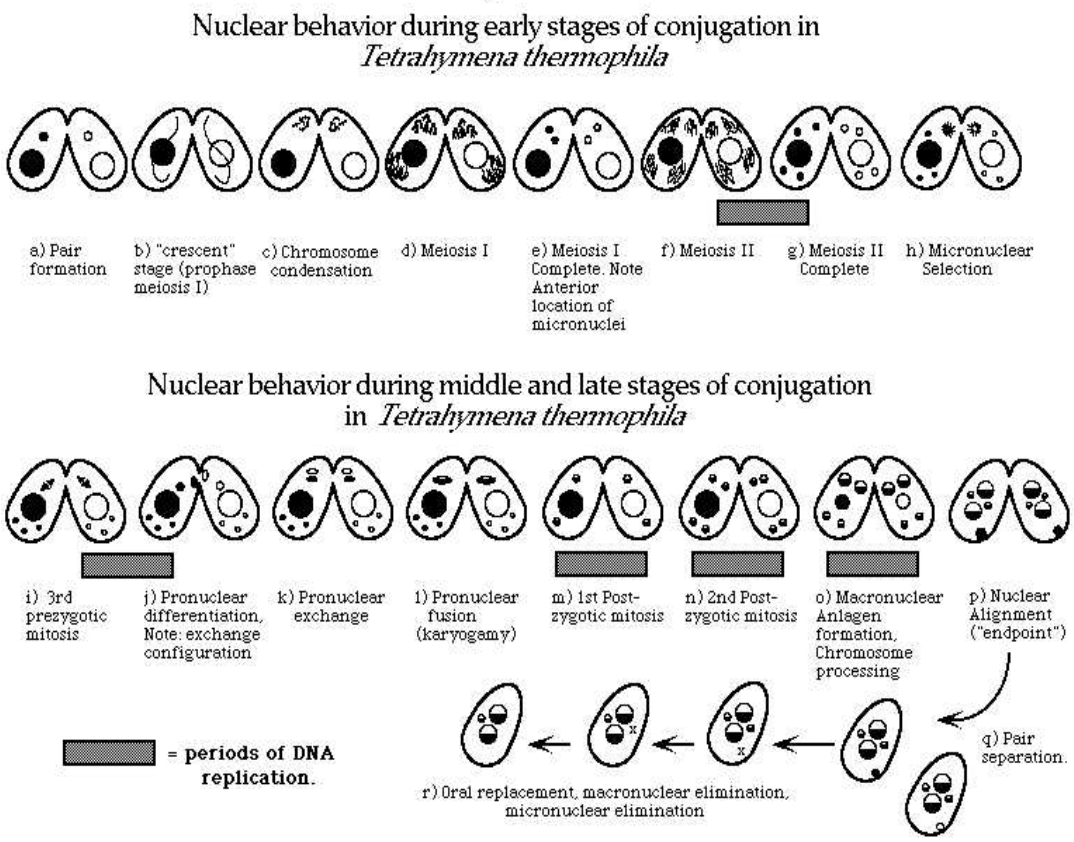


Figure 1.8 Nuclear events during conjugation of *Tetrahymena thermophila*.

into a 21 kb head-to-head palindrome. This rDNA minichromosome then gets amplified 5000-fold in the developing macronucleus in a single S phase (149).

The macronucleus lacks any structural features of mitosis such as kinetochores or spindle apparatus. Thus, in the vegetative cell cycle, it divides by amitosis. The macronucleus elongates, constricts and approximately half of the macronuclear DNA is distributed to each daughter cell. Inaccuracy could lead to the generation of progeny that contain deficiencies and imbalances that could possibly be lethal. However, strict regulatory mechanisms appear to exist to ensure that each macronuclear chromosome is maintained at 45C. Sometimes unusual cell cycles occur. In cases where the macronucleus has excess DNA, the cell divides without macronuclear DNA replication. Two rounds of macronuclear DNA replication can occur when there is less DNA in the macronucleus (145).

Transformation in *Tetrahymena*

Ciliates are unusual organisms since they are separated from the external medium not only by a plasma membrane, but also by an elaborate cell cortex composed of alveolar system of vesicles and a dense cytoskeleton known as epiplasm. Electroporation can be used to transform *Tetrahymena thermophila*, however the transformation efficiency is very low. In my thesis research, I have used a more recently developed method, called bioballistic or biolistic transfection to introduce circular plasmids into *Tetrahymena*. The biolistic gun (Biorad PDS-1000/He) has been used for this process (Figure 1.9). The transforming DNA is coated onto 10 μm gold particles and fired into starved vegetative or mating *Tetrahymena*. Transfection in the early phases of

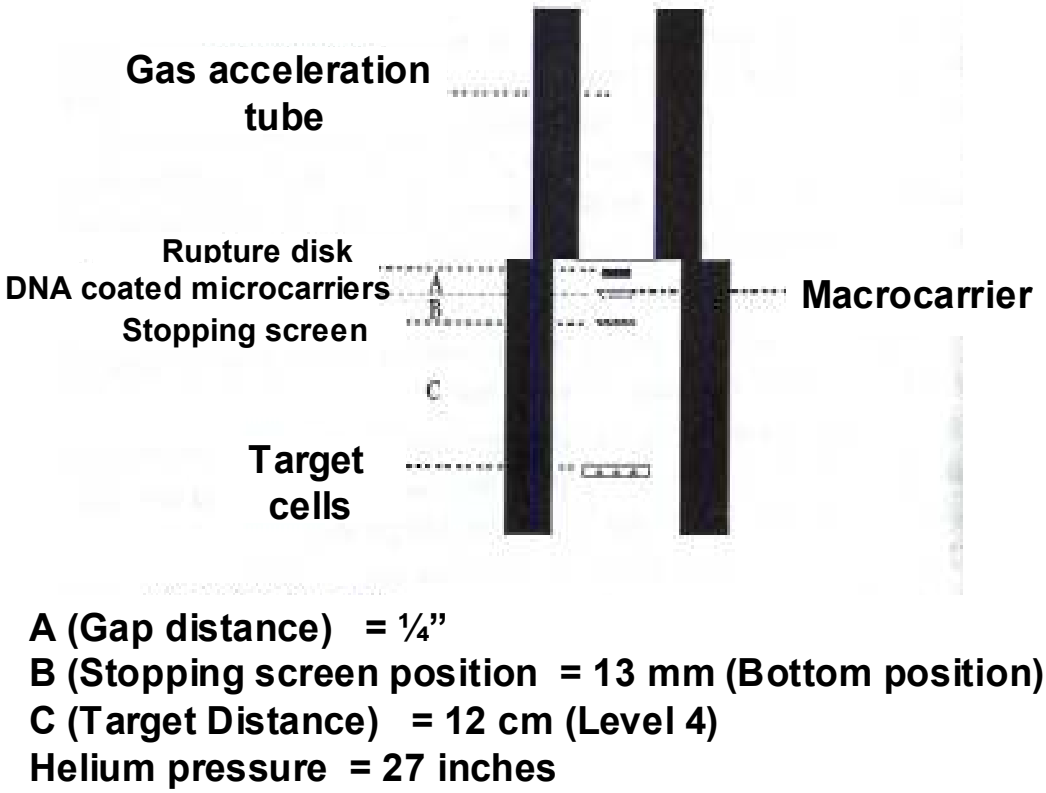


Figure 1.9 Biolistic gene gun (Biorad PDS-1000/He) used for transformation of *Tetrahymena thermophila*.

conjugation will introduce DNA into the micronucleus. Gene replacement in the micronucleus can be done by transformation of the micronucleus with linear DNA molecules which undergo homologous recombination. Micronuclear transformation does not result in altered phenotypes unless progeny macronuclei develop during conjugation. Transformation of mating cells in later stages of conjugation results in transformation of the developing macronucleus. Transgenes are propagated and expressed in vegetative cell cycle, but are not transmitted in the next sexual cycle (141).

In this study I set out to identify a new replication origin in *Tetrahymena thermophila* by isolating DNA segments that confer autonomous replication to plasmids. Because I was interested in studying the maintenance of episomes in the *Tetrahymena* macronucleus, all DNA constructs were transfected into the developing macronucleus of *Tetrahymena* at the anlagen stage. Later, I employed techniques like restriction mapping, subcloning and Southern blot to genetically dissect a newly isolated replicon. These studies have resulted in the identification of a DNA segment in *Tetrahymena* macronucleus, which is capable of autonomous replication in plasmids. Extensive genetic characterization revealed the presence of two replicons within this DNA segment, each of which can function individually as an ARS.

CHAPTER II

MATERIALS AND METHODS

Tetrahymena strains

Wild type *Tetrahymena* strains CU427 and CU428 were used for mating and transfection experiments. Cells were cultured in 2% PPYS + PSF to a density of 2 x 10^5 cells/ml, and starved in 10 mM Tris at the same density.

E. coli strains

EpiMax300 electrocompetent *E. coli* (Lucigen) were used for re-transformation experiments involving plasmids containing the pCC1FOS vector backbone. The vector is maintained in single copy when the cells are cultured in LB + chloramphenicol (12.5 μg/ml. Addition of the copy control induction solution (Epicenter) raised the plasmid to 50 copies per cell. The induction solution contains arabinose as an active ingredient which induced the *trf* gene in EpiMax300 *E. coli* cells. The *trf* gene product was necessary to induce the high copy number *oriV* in the pCC1FOS plasmid, thus generating more copies to increase plasmid yield.

Plasmids for transforming *Tetrahymena*

1) AN101: This pUC vector based plasmid contains a 10.5 kb segment that spans the entire micronuclear copy of the rDNA including chromosome breakage sequences (Cbs) elements that flank both ends of the macronuclear destined rDNA sequence. Upon transformation of *Tetrahymena*, the rDNA sequence is excised and rearranged into a 21 kb palindromic minichromosome. There is a

point mutation in the rDNA copy that confers resistance to paromomycin. The concentration of drug used to select for AN101 transformants was 100 μg/ml. In the transfection experiments, AN101 was used as a reporter plasmid for *Tetrahymena* cells that took up exogenous DNA.

2) TtMac # 29 pool: Plasmid was extracted from pooled colonies from *Tetrahymena* macronuclear sub-library # 29 and was used for transfecting *Tetrahymena*.
3) TtMac # 29 intergenic pool: Plasmid DNA was extracted from 4 clones in Library 29 which sequenced positive for intergenic regions in their inserts. This plasmid was used to transfect *Tetrahymena*.

Vectors used for sub-cloning

1) pSMART vector: pSMART vector (Lucigen) was used to blunt end clone the deletion fragments of TtARS1. The inserts (TtARS1 deletion fragments) were generated by specific restriction digests and had to be treated with DNA Terminator end repair kit (Lucigen) to create blunt ends. The blunt end inserts were ligated to pSMART vector.
2) pCC1FOS vector: A colony from TtMac library containing empty vector (without any *Tetrahymena* insert) was selected based on blue white selection and was used for cloning the deletion fragments of TtARS1.

Construction of the *Tetrahymena* macronuclear (TtMac) DNA library

A *Tetrahymena* macronuclear DNA library was manufactured by Lucigen by shearing the macronuclear genome and isolating fragments in the 5 to 7 kb range. These fragments were blunt-end cloned into pCC1FOS vector. pCC1FOS is an 8.1 kb plasmid that contains chloramphenicol resistance as selective marker and two *E. coli* origins of replication, an F-factor single-copy origin of replication and the inducible high-copy *oriV*. The *Tetrahymena* fragments were cloned into the Eco72*I* site of the vector, thereby disrupting a *lacZ* gene. Clones were selected by blue-white selection on Luria Bertani plates containing 12.5 μg/ml chloramphenicol and IPTG + X-gal. This *Tetrahymena* macronuclear library was divided into 31 sub-libraries, each representing 0.6X coverage of the genome (total genome size of *Tetrahymena* = 1 x 10^8 bp). The fraction of empty vectors in the library was reported to be below 2% by the manufacturer.

Screen for TtMac clones lacking the rDNA origin

rDNA is highly represented in the *Tetrahymena* macronucleus compared to non-rDNA chromosomes (9000 vs. 45C), and therefore steps were taken to eliminate plasmid clones that contained the rDNA origin. TtMac sub-library # 29 was selected at random and serial dilutions of the culture were plated onto LB + chloramphenicol (12.5μg/ml) plates at a density of 100 colonies per 100 mm petridish. The plates were incubated for 18-24 hours at 37^0C. Sets of ten colonies were pooled into LB + chloramphenicol (LB+ chl.) and allowed to grow for 8 hours in a 37^0C shaker. Plasmid DNA was extracted from these cultures using a Qiagen Spin Mini Prep Kit (Qiagen Inc.). The plasmid was eluted in 50 μl of sterile 10 mM Tris and used for PCR. In order to identify pools that

lacked the 1.9 kb rDNA origin in the plasmid samples, rDNA origin primers (forward primer: rDNA primer # 7 CTCGAGAAAAAATCAAAAAGAGATAAAAAGAC and reverse primer rDNA primer # 46 GTGGCTTCACACAAAATCTAAGCGC) were used. Only those sets that screened negative for the rDNA origin were further tested for replication competence in *Tetrahymena*. rDNA origin negative bacterial sets were combined to generate a pool of 100 colonies for transforming *Tetrahymena*.

Starvation of *Tetrahymena* strains

Wild type *Tetrahymena* strains (CU427 and CU428) were grown separately to a density of 2.5 X 10^5 cells/ml in 2% PPYS media (2% proteose peptone, 0.2% yeast extract, 0.1% sequestrene) containing PSF (100X solution of Antibiotic-Antimycotic containing penicillin, streptomycin and fungizone : Invitrogen). Cultures were grown at 30^0C in a shaker at a speed of 100 rpm. The flask size used was at least 10 times larger than the volume of the culture. The cells were harvested by centrifugation in GSA Sorvall ultracentrifuge at 3500 rpm for 5 minutes. Two washes were done with 10mM Tris without PSF by centrifuging at 3500 rpm for 5 minutes each. The cells were resuspended in 10 mM Tris containing PSF at a density of 2.5 X 10^5cells/ml and incubated in shaker for 18 hours for starvation. After 18 hours, 25 mls of each starved strain was added into a T150 tissue culture flask for mating at 30^0C under stationary conditions.

Biolistic transfection of *Tetrahymena*

All DNA constructs were transfected at the macronuclear anlagen stage into *Tetrahymena* mating pairs of CU427 and CU428 strains. The developmental cycle of the

mating pairs was monitored by staining an aliquot of the cells with acridine orange and observing under fluorescence microscope. The anlagen stage was determined to occur approximately 8 hours and 40 minutes post-mating. At this stage, the mating cells were harvested by centrifugation at 3000 rpm for 2 minutes and resuspended in 1 ml of 10 mM HEPES (pH 7.5). The cells were transferred onto a sterile Whatman filter paper in a 100-mm-diameter petridish. The Whatman paper was presoaked with 1 ml of 10 mM Tris (pH 7.5). The petridish was then placed in the Biorad gene gun chamber. The macrocarrier used for transfection was 10 μm gold particles (Biorad) which were coated with DNA mixture containing 35 μg AN101 plasmid and 35 μg of the experimental plasmid (TtMac library DNA or the deletion constructs). The DNA was then introduced into *Tetrahymena* cells as per manufacturer's instructions (Biorad). After transformation, the cells were transferred into 500 ml flasks containing 50 mls of 10 mM Tris and incubated under stationary conditions at 30°C for 18 hrs for completion of the development cycle. The cells were then re-fed with 5%PPYS to a final concentration of 2% and incubated in shaker for 6 hrs (149).

Plating of Tetrahymena *transformants*

After refeeding, paromomycin was added to the transformed culture at a final concentration of 100μg/ml. The cells were then diluted in 2%PPYS + PSF + Paromomycin, and subsequently plated into 96 well culture plates (Corning & Falcon). Two plates each of undiluted, 1:10 and 1:100 dilution of each transformation were made and the plates were incubated in a dark and moist chamber at 30^0C for 3-4 days.

En masse culturing

After plating the transformants into 96 well plates, the remaining transformed culture was passaged as en masse transformants for population studies in 25 mls of 2% PPYS + PSF + paromomycin (100µg/ml) until seven passages. The en masse transformants culture was used for plasmid re-isolation assay and for Southern blot analysis to study plasmid maintenance.

Plasmid extraction from *Tetrahymena* and re-transformation into *E. coli*

50 mls culture of *Tetrahymena* en masse transformants was grown to a density of 2.5 X 10^5 cells / ml and harvested by centrifugation at 3500 rpm for 5 minutes. The pellet was washed with 10 mM Tris and plasmid was extracted by using a Qiagen Midiprep Kit. The plasmid was eluted in 40 µl of 1x TE (10 mM Tris, 1 mM EDTA). Dpn*I* digestion of the extracted plasmid was done to assure that any unreplicated input DNA was eliminated. 2 µl of digested DNA was electroporated into 20 µl of EpiMax 300 cells (Epicentre). The transformants were plated onto LB + chl. agar plates and incubated for 18-24 hrs at 37^0C. Plasmid DNA was extracted from the *E. coli* transformants using a Qiagen Spin Mini prep Kit, followed by specific restriction digestion of the plasmid.

Genomic DNA isolation and Southern blot analysis of transformants

25 ml cultures were used from each en masse transformant cell lines to prepare total genomic DNA using standard protocol, wherein the cells were harvested by centrifugation at 3500 rpm for 5 minutes followed by a wash with 10 mM Tris. The pellet was resuspended in 2.5 mls of 10 mM Tris and then alkaline lysis was carried out

with 2.5 mls of NDS (pH = 9.5, pre-heated at 55^0C) and 25 µl of 20 mg/ml solution of Proteinase K (Sigma). NDS consists of 90 mls 0.5 M EDTA (pH=9.5), 1 ml 1M Tris (pH=9.5) and 10 mls of 20% sodium dodecyl sulfate in a total volume of 100 mls of distilled water. The samples were then incubated at 55^0C for 4 hrs to overnight. NDS lysis was followed by phenol chloroform extraction and ethanol precipitation. The DNA was quantified using a spectrophotometer.

50 µg of total genomic DNA was used from each sample for appropriate restriction digestion. For each sample, digest was set up with 3µl 10X buffer, 1.5 µl enzyme and H_20 for a 30µl reaction and the tubes were incubated in 37^0C waterbath for 4 hrs. Digested DNA was run on 0.6% agarose gels. DNA in the gel was then depurinated with 0.1 M HCl for 15 minutes twice to facilitate large fragments that are difficult to transfer to get blotted. The DNA was then denatured by soaking it in 0.4M NaOH for 15 minutes twice, which separated the double-stranded DNA into single-stranded DNA. Vacuum Blotter (Biorad) was used to transfer the DNA onto charged nitrocellulose membrane (Genescreen Plus). Radiolabelled probes were prepared using Megaprime DNA labeling Kit (Amersham Biosciences). The blot was crosslinked using UV Stratalinker (Biorad) and prehybridized with Church solution at 50^0C for 3 hrs to overnight (100 mls of Church solution: 30 ml of 20% SDS, 50 ml of 1M Na_2HPO_4 200 µl of 0.5 M EDTA, 2 mls of 5 mg/ml salmon sperm DNA, 1 ml of 100 X Denhardt's solution, H_2O, store at 4^0C. 100 X Denhardt's solution: for 500 ml: 10 g Ficoll 400; 10 g polyvinylpyrrolidone, 10 g BSA, H_2O store at -20 °C). Hybridization to the probe was done overnight and the blot was washed

at 50^0C twice for 15 minutes with 0.1M Na_2HPO_4 + 0.1% SDS and exposed to a phosphorimager screen.

General cloning strategy to produce deletion constructs of TtARS1

Deletion constructs of TtARS1 was made primarily by two methods, restriction digestion and re-ligation, as well as by inverse PCR. A description of the first set of deletion derivatives of TtARS1 is shown in Figure 2.1. The ΔNhe-1 construct of TtARS1 was made by NheI digest followed by self-ligation of the larger fragment. This contained the 8.1 kb vector and 4.0 kb of insert. The 2.8 kb internally deleted fragment was blunt end cloned into the pSMART vector. NsiI digestion of full length TtARS1 resulted in the release of a 1.9 kb fragment TtARS1 insert. The larger fragment after re-ligation was called ΔNsi-1. The deleted fragment was again cloned into pSMART as before. These constructs in pSMART were called Nhe-2 and Nsi-2 respectively. NheI and Sph*I* double digestion of Nhe-1 resulted in an internal deletion of 3 kb segment from the insert. Self-ligation of the larger fragment produced Nhe-L deletion construct that contained only the leftmost 1 kb region of TtARS1 insert in pCC1FOS vector backbone. The 3 kb deleted fragment was blunt end cloned into pSMART vector and was named Nhe-R1. The Nhe-R1 insert fragment contains a predicted intergenic region of 1.2 kb. This region was PCR amplified and cloned into pSMART vector backbone. The construct was designated Nhe-R2.

The second set of deletion constructs of TtARS1 was derived by using Nhe-L construct as template (Figure 2.2). Deletion derivatives C6, L6, R6, L8 and R8 were constructed by PCR amplification of the respective inserts using TtARS1 as template

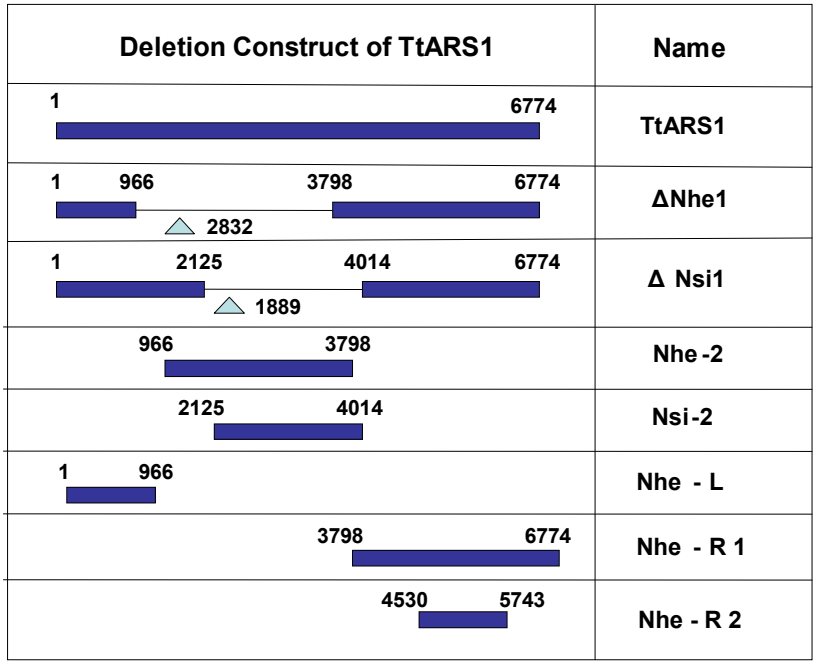


Figure 2.1 Deletion constructs of TtARS1

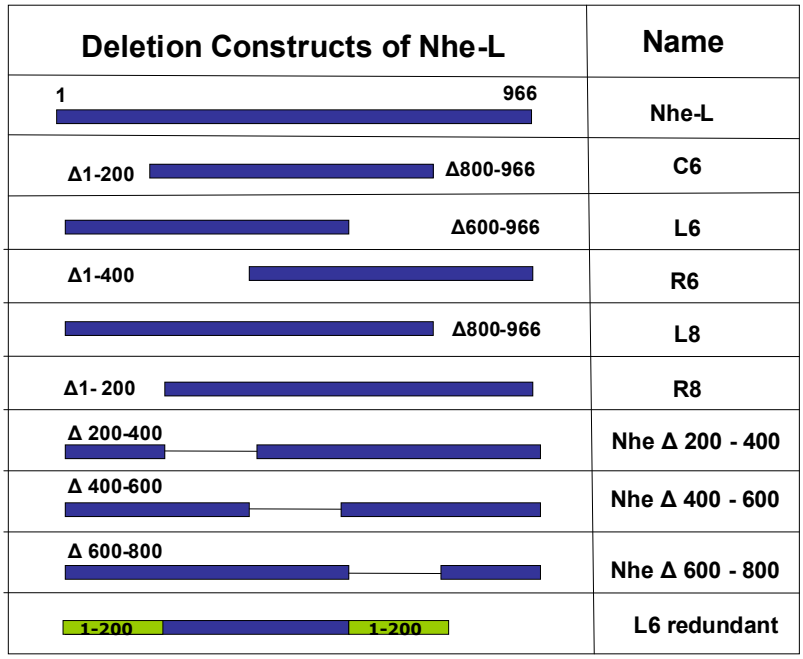


Figure 2.2 Deletion constructs of Nhe-L

and Pfu Turbo polymerase. The inserts were cloned into pCC1FOS vector at BamHI site. Inverse PCR followed by self-ligation was used to generate Nhe Δ 200 – 400, Nhe Δ 400 – 600 and Nhe Δ 600 – 800 constructs. The construct named L6 redundant was made by PCR amplification of the leftmost 200 bp segment of L6 and ligating it to the opposite end of L6 deletion construct. A third set of deletion derivatives (Figure 2.3) of TtARS1 was made by using L8 as the template. LΔ1-100, LΔ50-150, LΔ100-200, LΔ600-700, LΔ650-750 and LΔ700-800 were generated by inverse PCR using Pfu Turbo polymerase. The PCR products were self-ligated after kinasing the fragments.

Analysis of intergenic regions from TtMac sub-library # 29

Plasmid from 28 individual clones from TtMac sub-library # 29 was sequenced to obtain clones that contained predicted intergenic regions in their inserts. Gene prediction on the inserts predicted 5 intergenic regions. The size of these intergenic segments and the macronuclear scaffold that contain them are shown in Table 2.1. Plasmid was extracted from these clones and pooled together for transfection into *Tetrahymena*.

Cloning of rDNA 5'NTS into TtARS1 deletion constructs

Two tandem copies of the rDNA 5' NTS which contains the rDNA origin of replication were cloned into the HinDIII site of C6 and into the BamHI site of Nhe-L deletion constructs of TtARS1. They were named Chimera 1 and Chimera 2 respectively. These plasmids were transfected into *Tetrahymena* by biolistic transfection.

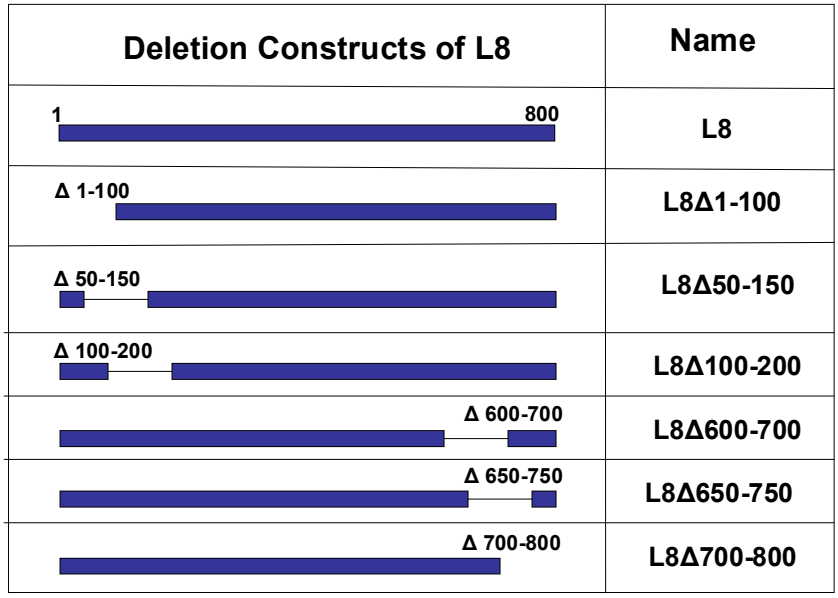


Figure 2.2 Deletion derivatives of L8

Table 2.1 Size of predicted five intergenic regions in TtMac sub-library # 29 clones

Macronuclear Scaffold	Size of intergenic region
8254584	1734
8254378	613
8254764	3546
8253817	340
8254811	709

CHAPTER III

RESULTS

In the introduction chapter I stated that the only origin of DNA replication in *Tetrahymena thermophila* that has been identified and characterized in detail is the one present in the ribosomal DNA minichromosome. The regulation of the rDNA minichromosome is different from that of the bulk non-rDNA chromosomes of the macronucleus in that it undergoes locus specific gene amplification. During subsequent vegetative divisions, replication of the rDNA and other macronuclear chromosomes is under cell cycle control. rDNA replication initiates within a 1.9 kb segment in the 5' non-transcribing spacer (NTS) region (58, 59, 60). However, in spite of the detailed knowledge of the cis-acting determinants of the rDNA origin, the mechanism of locus specific amplification remains elusive. No determinants have been identified in the rDNA thus far that are amplification specific. Moreover, the detailed knowledge of only one replicon in *Tetrahymena* limits the possibility for comparative studies, and hence I set out to isolate additional *Tetrahymena* replicons. The approach that I used in my thesis research was to isolate non-rDNA fragments that support autonomous replication.

A commercially prepared *Tetrahymena* macronuclear DNA library (see chapter II) was used as my source of plasmid DNA. The size of the inserts in the library was kept in the lower range of 5 to 7 kb because of the extremely high AT content of *Tetrahymena* DNA and the instability of large AT-rich inserts in *E. coli*. This library was the best one available at the time and was considered to be sufficient for the isolation of

non-rDNA replicons, considering the size of the *Tetrahymena* rDNA replicon as well as other eukaryotic replicons. The first objective of my thesis research was to identify non-rDNA segments that function as replication origins in *Tetrahymena* macronucleus. I chose to achieve this by isolating sequences that confer autonomous replication to plasmids. The search for DNA fragments that function as replicators in *Tetrahymena* involved three steps, introduction of plasmid DNA from a macronuclear library into *Tetrahymena*, propagation of *Tetrahymena* transformants to permit several rounds of DNA replication, and re-isolation of plasmid DNA from *Tetrahymena* and re-transformation into *E. coli.* This is called the ARS isolation strategy.

ARS isolation strategy

The basic strategy for identifying of autonomous replicating sequences in *Tetrahymena* is shown in Figure 3.1. The library was constructed by shearing the macronuclear genome to 5 to 7 kb range fragments and cloning the fragments into pCC1FOS vector. The vector has an *E. coli* origin of replication. Plasmid DNA from the library was transfected into *Tetrahymena*. The transformants were propagated for several generations and plasmid DNA was subsequently shuttled from *Tetrahymena* transformants into *E. coli* after the removal of residual input DNA. The input DNA was removed by Dpn*I* digestion. Thus the autonomous replicating sequences were isolated from *Tetrahymena*. The different steps in this strategy are explained in detail below.

Starting material - plasmid DNA containing potential replication origins

The *Tetrahymena* DNA library that was generated was cloned as 31 sub-

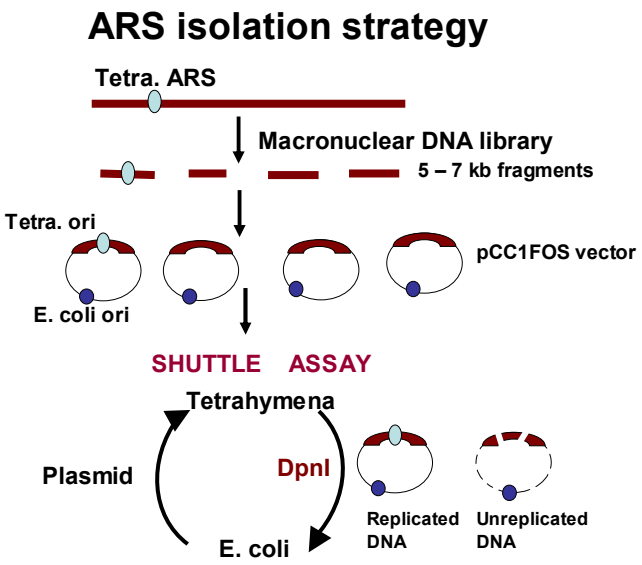


Figure 3. 1 ARS (Autonomous replicating sequence) isolation strategy to identify non-rDNA replicons in *Tetrahymena*.

Tetrahymena macronuclear DNA library was constructed by shearing of the macronuclear genome and size fractionating the fragments to 5-7 kb length. The fragments were cloned into pCC1FOS vector, which has an *E. coli* origin of replication. Plasmid pool from the library was transfected into *Tetrahymena* and the transformants were propagated to few generations. Plasmid extracted from the transformants was shuttled into *E. coli* after Dpn*I* digestion to remove any unreplicated input DNA.

libraries, from which sub-library # 29 was selected at random. Plasmid DNA from a set of 100 library clones (from sub-library #29) that screened negative for the rDNA replication origin was used as the starting material for transforming *Tetrahymena*. This rDNA pre-screening process, depicted in figure 3.2, consisted of PCR amplification of plasmid DNA from pools of ten library clones using primers that were specific for the rDNA origin (1.9 kb 5' NTS). The PCR amplified product was electrophoresed in a 1% agarose gel. This screening step was essential since the rDNA is very highly represented in the macronucleus compared to the non-rDNA chromosomes (9000C vs. 45C). As shown in figure 3.2, positive PCR reactions were observed in pools 1-6, which was indicative of contamination with cloned rDNA origins. Approximately 25% of the pools screened negative for rDNA origin. Ten pools that lacked the rDNA origin were combined to form the pool of 100 library clones. The next step was to introduce these plasmids into the developing macronucleus of *Tetrahymena* to identify that support autonomous replication.

Controls for maximum transfection efficiency of Tetrahymena*:* The two methods available for transforming of *Tetrahymena* are electroporation and biolistic transfection. Biolistic transfection can be done using two types of adaptors, the mono- adaptor and the hepta adaptor. The difference between the two is that the mono-adaptor uses a single channel for flow of pressure whereas the hepta adaptor has seven channels through which air can flow. Hence, the hepta adaptor covers more surface area on the petridish which contains the cells that are to be transformed. A comparison of transfection efficiencies between electroporation and biolistic transfection was done using AN101

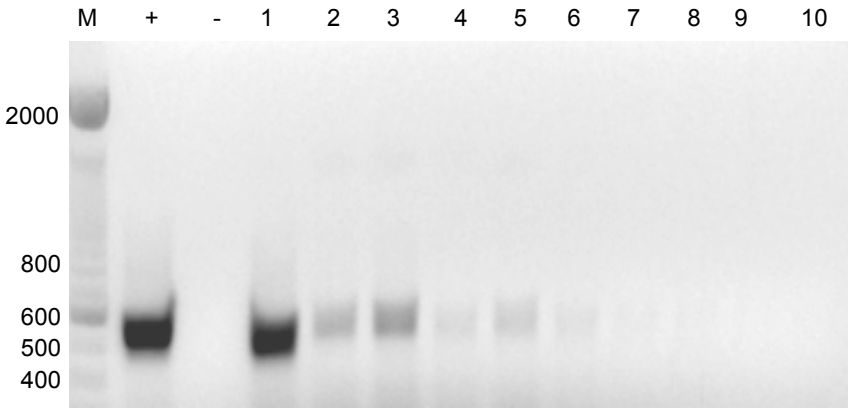


Figure 3.2 Screening of TtMac sub-library # 29 for clones lacking rDNA 5' NTS.

Plasmid from 10 sets of 10 TtMac clones were used for PCR amplification using rDNA origin (5' NTS) specific primers. M designates 100 bp marker, + is rDNA origin containing plasmid used as a positive control for PCR, and - is negative PCR control. 1 – 10 designates PCR product derived from TtMac library plasmid DNA pools of 10.

plasmid DNA. The AN101 plasmid contains a copy of the *Tetrahymena* rDNA in a pUC vector backbone. In case of electroporation and biolistic transfection with mono-adaptor, 10 μg of AN101 DNA was used, whereas for the hepta adaptor 70 μg of DNA was transfected. The transformed culture was plated into 96 well plates at different dilutions. The plates were scored and the average number of transformants obtained per μg of transformed DNA was calculated. The transfection efficiencies were 480 transformants / μg of transformed DNA in case of electroporation, 1520 for mono-shot biolistic transfection and 4160 for hepta shot biolistic transfection. It was observed that the efficiency of transfection was more in case of biolistic transfection as compared to electroporation. Moreover, the hepta shot was at least three times more efficient than the single shot method. Therefore, biolistic transfection with the hepta adaptor was used as the gene delivery system for introducing DNA fragments with potential origin function.

One of the key steps in this study involved the re-isolation of plasmid DNA from *Tetrahymena* and re-transformation of *E. coli*. There exists a protocol for isolation of total genomic DNA from *Tetrahymena*, however, there were no established methods for plasmid isolation. One idea was to isolate total genomic DNA (chromosomal DNA + plasmid DNA) from *Tetrahymena* transformants and use this to transform *E. coli*. For this purpose, a control experiment was done to examine the effect of bulk genomic DNA contamination on electroporation of *E. coli* with episomal DNA. A mock experiment was done where varying amounts of plasmid DNA were added to a fixed amount of *Tetrahymena* genomic DNA. The amounts of genomic DNA relative to plasmid DNA as well as the total number of *E. coli* transformants obtained is shown in Table 3.1. The

plasmid used for this experiment was pUC119, and hence the *E. coli* transformants were plated on LB agar containing ampicillin (100μg/ml).

The above experiment revealed that the rate-limiting factor that determined the transformation efficiency is the amount of plasmid DNA that was added to the genomic DNA. The amount of genomic DNA did not affect the transformation frequency in an adverse way. This result was very useful since it provided a way in which episomes that replicated in *Tetrahymena* could be isolated and transformed into *E. coli* successfully.

Controls for re-isolating plasmid DNA from Tetrahymena: The above experiment revealed that transformation of plasmid DNA together with genomic DNA into *E. coli* does not significantly decrease transformation frequency. However, this method has a disadvantage. The objective of this assay is to isolate autonomously replicating DNA molecules from *Tetrahymena*. But, in situations where the plasmid replicated poorly in *Tetrahymena* or the transforming pool of DNA (from TtMac library) contained few functional replicons, the total genomic DNA extracted will contain a very low ratio of plasmid DNA to chromosomal DNA. This can be an issue since large amounts of *Tetrahymena* genomic DNA cannot be electroporated into *E. coli*. Thus the overall *E. coli* transformation efficiency might be too low for plasmid re-isolation. A way to resolve this problem is to extract plasmid from *Tetrahymena*. The alkaline lysis protocol that is developed for plasmid isolation from bacteria, has also been used for isolation of episomes from mammalian and plant cells. Although, it had never been tried with ciliates, logically it should be applicable.

Table 3. 1 Optimization of DNA concentration for re-transformation of *E. coli* (electroporation) with DNA extracted from *Tetrahymena*.

Mixing experiments of *Tetrahymena* genomic DNA with exogenous plasmid DNA was done to study the effect of genomic DNA concentration on electroporation of *E. coli*. The ratio of plasmid:genomic DNA and the total number of *E. coli* transformants obtained have been shown.

Plasmid DNA concentration (ng)	Genomic DNA concentration (ng)	Plasmid : genomic ratio	Average # E. coli transformants
10	0.0	10:0	270
1	0.0	1:0	95
0.1	0.0	0.1:0	18
10	100	1:10	216
1	100	1:100	30
0.1	100	0.1:100	19
0.0	100	0:100	0.0

With this rationale, a mock experiment was designed in which a Qiagen plasmid extraction Mini column was used. A 7 ml log phase culture of wild type *Tetrahymena* (strain CU428) was grown to a density of 2.5 x 10^5 cells/ml. The cells were centrifuged and the pellet was resuspended in the buffer provided in the kit. However, immediately after the addition of lysis solution, 500 ng of exogenous plasmid DNA was added to the cells. The plasmid used contained two copies of the rDNA origin (5' NTS) within the pUC 118 vector. This particular plasmid was used because of the availability in the lab, rather than for any technical reason. The Qiagen protocol was continued to completion and in the final step the DNA was eluted in 40 µl of sterile distilled water. 20 µl of the eluted DNA was electrophoresed in a 1% agarose gel. The rest of the sample (20 µl) was digested with BamHI to test the integrity of the plasmid. The controls that were run on the gel were uncut input DNA and BamHI digested input DNA (pUC 2 x 1.9). It was observed that the recovery of plasmid in the eluted fraction was approximately 10% (data not shown). Most of the plasmid that was retrieved was supercoiled and the integrity of the plasmid was confirmed by digestion (data not shown). Most importantly, the eluted fraction contained only plasmid DNA and no contaminating *Tetrahymena* DNA or RNA as observed by ethidium bromide staining.

Thus, from the above experiments, transfection of DNA into *Tetrahymena* and a method for re-isolating plasmid from the *Tetrahymena* transformants was standardized. At this point, an attempt at the actual experiment was done. Since, the goal was to isolate new *Tetrahymena* replicons, the idea was to transfect *Tetrahymena* with a macronuclear DNA library (TtMac library) to search for elements that can function as replicators. 35

μg of plasmid DNA from a pool of 100 clones from the TtMac library that lacked the rDNA 5' NTS was transformed into *Tetrahymena* together with 35 μg of AN101 plasmid as the co-transforming vector. AN101 is a pUC118 vector-based plasmid that contains a copy of the entire micronuclear rDNA of *Tetrahymena*, including the chromosome breakage sequence (Cbs). This rDNA copy harbors a point mutation that confers resistance to the antibiotic paromomycin. Once transformed into *Tetrahymena*, AN101 undergoes rearrangement by breakage at Cbs elements and palindrome formation, thereby generating a linear molecule. The plasmid vector sequences are removed. Therefore, these linear molecules cannot be re-isolated from *Tetrahymena* by re-transforming *E. coli*. Since AN101 was used as co-transforming plasmid, paromomycin at a final concentration of 100μg/ml was used to select for the cells that took up DNA.

Tetrahymena cells were transformed biolistically with the hepta adaptor at the macronuclear anlagen stage of development. Macronuclear anlagen formation is the stage where the new macronucleus develops from one of the newly generated micronuclei. A key stage in the developmental cycle of *Tetrahymena* includes genome-wide endoreplication, when the copy number of the non-rDNA chromosomes increase from 2C to 45C, and the rDNA amplifies from 2C to 10,000C. As opposed to this, during vegetative replication, every DNA sequence only replicates once per cell cycle. In this study, the maintenance of autonomously replicating plasmids was used as the primary criteria to identify functional replicons. For this purpose, the plasmid library pool has to be introduced into the developing macronucleus of mating cells, so that any

introduced replicon will have the opportunity to endoreplicate to 45C. After mating cells complete development, they are re-fed with fresh PPYS media, and vegetative DNA replication and copy number control mechanism start actively. During this phase, if a insert in the TtMac library DNA pool (that has been introduced into *Tetrahymena*) contains a *Tetrahymena* origin of replication, then this plasmid should replicate in *Tetrahymena* and be maintained as an episome. Copy number control mechanism will ideally take over and establish a fixed number of plasmids per *Tetrahymena* cell.

A 6.7 kb DNA fragment supports autonomous replication in *Tetrahymena*

Mating cultures of wild type *Tetrahymena* strains CU427 and CU428 cells were subjected to biolistic transfection with the pool of non-rDNA fragments from 100 TtMac clones (5-7 kb inserts). Anlagen stage occurred 8 hours and 40 minutes post mating as observed by staining with acridine orange. After transfection, the cells were allowed to complete development and then re-fed the next day with 5%PPYS + PSF to a final concentration of 2%PPYS and selected with 100μg/ml paromomycin six hours after re-feeding. A portion of the transformation cultures were plated into 96 well microtiter dishes. The dilutions of the transformed culture that were plated include two plates each of undiluted, 1:10 dilution and 1:100 dilution. The rest of the transformed cells was propagated as an “en masse” culture. The 96 well plates were used to establish clonal lines of the transformants, whereas the en masse cultures were used to study the maintenance of the plasmid *in vivo*.

Plasmid extraction from en masse culture and E. coli re-transformation

The *Tetrahymena* transformants were propagated en masse in 2%PPYS + paromomycin (100µg/ml) to kill untransformed cells and plasmid DNA was extracted from paromomycin resistant transformants at passages 1 and 3 (approximately 10 and 25 fissions, respectively). Plasmid extraction was done from 50 mls of *Tetrahymena* transformants at a density of 2.5 x 10^5 cells/ml with a Qiagen Midi prep column. The plasmid was then treated with Dpn*I* restriction endonuclease to remove any residual input plasmid. Dpn*I* digests fully methylated DNA. *Tetrahymena* has very low efficiency of DNA methylation and hence any DNA that has replicated in *Tetrahymena* will be resistant to Dpn*I*. The input DNA for transfection was obtained from *E. coli* cells that contained a functional deoxyadenosine methylase and hence any residual input DNA will be digested by Dpn*I*. Dpn*I*-resistant plasmid DNA was then re-transformed into *E. coli* by electroporation to recover plasmids that replicated in *Tetrahymena*. A total of twelve *E. coli* transformants were obtained. This was the first indication that the DNA pool that was transfected into *Tetrahymena* probably had some potential replicons.

There are two BamHI sites that flank the insertion site in the library vector backbone, pCC1FOS. Plasmid was extracted from the *E. coli* re-transformants and digested with BamHI to release the insert from the vector (Figure 3.3). The BamHI digested products were electrophoresed in a 1% agarose gel. Only two out of the twelve *E. coli* transformants contained an insert (lanes 1 and 2 in figure 3.3). The rest contained only vector sequences (data not shown). This extremely high rate of loss of plasmid integrity may be due to deletions associated with the *E. coli* re-transformation process.

Plasmid from the two clones that contained inserts was sequenced with vector-specific primers that flank the insertion site. The DNA sequence revealed that the two clones were identical. The insert sequence was used to query the *Tetrahymena* Genome Database (150) using BLAST. The fragment was present on a 405 kb contig. The contig number was 8254464 and the sequence corresponded to positions 264443 to 265516. The macronuclear scaffold was later designated by TIGR as CH445556. The insert was named TtARS1 (*Tetrahymena thermophila* autonomous replicating sequence 1) and the total size was 6,774 basepairs.

Thus, a putative sequence in the *Tetrahymena* macronucleus that supports autonomous DNA replication was identified. The next step was to verify that TtARS1 functions as an origin and to assess plasmid stability during long term propagation in *Tetrahymena*. For this purpose, TtARS1 was co-transfected into *Tetrahymena* with AN101 plasmid, and the transformants were analyzed in both en masse culture as well as in 96 well microtiter dishes. The en masse population was passaged over 60 generations. Since, AN101 plasmid which confers paromomycin resistance was used as the co-transforming vector, the transformed culture contains a mixture of two populations. One population should consist of cells that contained only AN101 and hence are resistant to paromomycin. The second transformant population should harbor both AN101 and TtARS1.

TtARS1 replicates autonomously for 60 fissions in Tetrahymena*:* Southern blot analysis was used to study the maintenance of TtARS1 as an episome. Each passage of *Tetrahymena* equals approximately 7 fissions, with the first passage consisting about

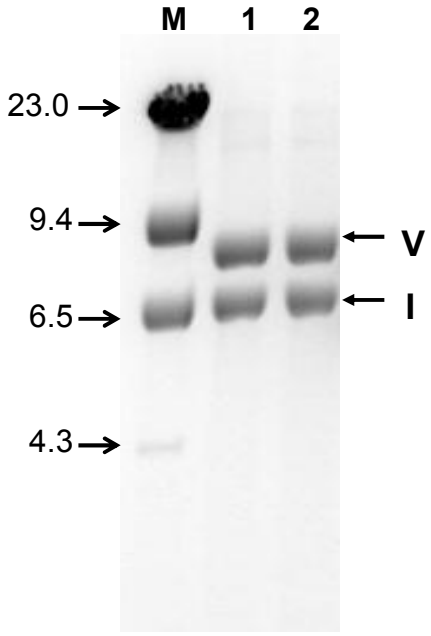


Figure 3.3 Initial isolation of TtARS1 by shuttle assay.

Plasmid extracted from *Tetrahymena* transformants of TtMac library DNA was transformed into *E. coli* after Dpn*I* digestion. Plasmid DNA from *E. coli* re-transformants was treated with BamHI enzyme, which released the 6.8 kb insert (I) from 8.1 kb vector backbone (V). M designates Lamda HinDIII marker, 1 and 2 represent BamHI digested plasmid from two *E. coli* re-transformants.

15 fissions. Southern blot was done with total genomic DNA extracted at 15, 30, 45 and 60 fissions of *Tetrahymena* transformants. 50 μg of genomic DNA was digested with BamHI enzyme and run on a 0.4% agarose gel. BamHI digestion releases a 15 kb chromosomal fragment and a 6.8 kb episomal fragment from *Tetrahymena* transformants containing TtARS1 (Figure 3.4). DNA was transferred to a Genescreen plus membrane and probed with the radiolabelled TtARS1 insert fragment (Figure 3.4). Southern blot result showed the presence of the episomal signal in addition to the endogenous chromosomal signal in the genomic DNA from transformants (figure 3.4). Genomic DNA from untransformed *Tetrahymena* cells produced a single band (endogenous) in Southern blot. Therefore, it was concluded that TtARS1 is maintained as an episome in *Tetrahymena* transformants for 60+ fissions. The copy number of the episome is remarkably similar to that of the endogenous chromosomal copy.

The above result confirmed that the 6.8 kb TtARS1 segment has a *Tetrahymena* origin of DNA replication. Once introduced into the cells as a plasmid, it can replicate using its origin and the plasmid copy number stabilizes to a level that is analogous to the endogenous chromosomal copy. As a control the pCC1FOS vector was co-transfected into *Tetrahymena* with AN101. The vector could not be maintained on its own in *Tetrahymena* as observed by Southern blot analysis (data not shown). Also, plasmid extraction followed by re-transformation into *E. coli* produced no colonies, indicating that the empty vector does not support autonomous replication in *Tetrahymena*.

Analysis of clonal TtARS1 transformants: As stated before, a portion of the *Tetrahymena* TtARS1 transformants were plated into 96 well microtiter dishes to obtain

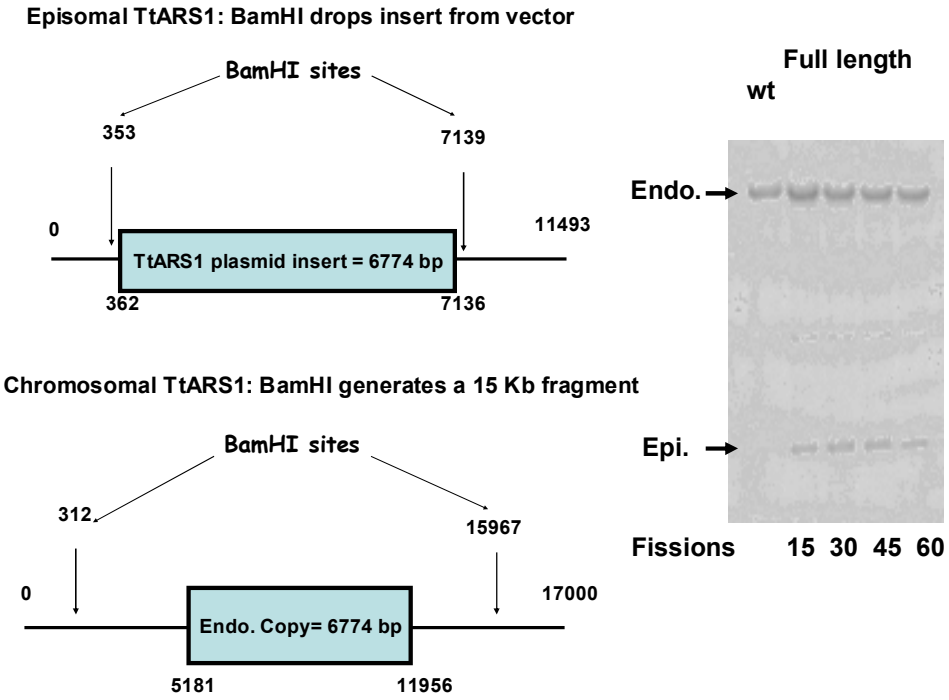


Figure 3.4 TtARS1 is maintained as an episome in *Tetrahymena*.

eft panel depicts the different size fragments produced on BamHI digestion of genomic DNA from TtARS1 transformants. BamHI digestion releases the 6.7 kb TtARS1 insert from the 8.1 kb vector, and it generates a 15 kb fragment in case of the endogenous chromosomal locus. Panel on the right shows Southern blot data using total genomic DNA from TtARS1 transformants extracted at the mentioned fissions. TtARS1 insert fragment was used as radiolabelled probe. Endo. represents the signal from endogenous native chromosome and epi. represents the signal from the episome. wt represents DNA from untransformed *Tetrahymena* cells which show the presence of endogenous signal only.

clonal lines. The 96 well plates that contained 1:100 dilution of the transformed culture had on an average twelve to sixteen paromomycin resistant wells. The higher dilution was used in order to maximize the chance of clonal lines based on the Poisson distribution. Cultures from a total of thirty wells from the 1:100 dilution plate were grown to screen for TtARS1 co-transformants by PCR amplification of the vector backbone. This was done primarily to determine the co-transfection frequency, which means the number of cells that contained both AN101 plasmid as well as TtARS1. Had there been a reporter in TtARS1 plasmid, this would not have been necessary. However, the lack of any selectable marker in TtARS1 compels the use of PCR technique to look for co-transformants. The primers used for PCR were specific for the pCC1FOS vector (forward primer: GTG CTG CAA GGC GAT TAA GTT GG, reverse primer: TGA CCA TGA TTA CGC CAA GCT ATT TAG G).

Genomic DNA was prepared from 1 ml cultures of transformants grown in 2% PPYS containing paromomycin (100 μg/ml) to a density of 2.5 X 10^5 cells/ml. This was done at passage 1 of the transformants. The DNA was digested with Dpn*I* to remove residual input DNA and 100 ng of genomic DNA per sample was used for PCR using Taq DNA polymerase. Sixteen of the thirty wells analyzed showed the presence of TtARS1 (data not shown). Thus, the co-transfection efficiency in this experiment was approximately 50%. Co-transformants were propagated until the 10^{th} passage which corresponds to over 85 fissions in *Tetrahymena*. Based on the Poisson distribution, the majority of these lines should be clonal. Total genomic DNA was isolated from these cultures, followed by PCR amplification with pCC1FOS vector specific primers. The

PCR product was electrophoresed in a 1% agarose gel and is shown in Figure 3.5 (lanes 1 to 16). PCR data showed that about 50% of the presumably clonal lines maintained TtARS1. This indicated that TtARS1 can stably propagate in *Tetrahymena* for over 85 fissions. Thus, at this point I conclude that this non-rDNA replicon of *Tetrahymena*, TtARS1, is capable of long term propagation as an episome. Moreover, a plasmid maintenance assay has been developed for the model system *Tetrahymena*, which wasn't available before this study.

Deletion analysis of TtARS1

Deletion of an internal fragment of 2.8 kb in TtARS1 does not affect autonomous replication

Having identified a non-rDNA replicon in *Tetrahymena*, the next goal was to delineate the origin in order to find the functional determinants of the replicon. This was first done by deletion analysis using restriction enzymes. The upper panel of figure 3.6 shows a graphical representation of the first two deletion fragments to be studied. These constructs, designated ΔNsi-1 (5.0 kb insert) and ΔNhe-1 (4.0 kb insert), contained internal deletions of 1.9 kb and 2.8 kb respectively. *Tetrahymena* was co-transformed with AN101 and each of the deletion constructs and paromomycin-resistant transformants were selected. Southern blot of en masse cultures of transformants was used to study the maintenance of these constructs in *Tetrahymena*. The transfection protocol used was the same as previously described (see methods chapter).

Southern blot analysis with en masse culture: After transfection, a portion of the culture was plated into 96 well microtiter dishes to monitor transfection frequency, and

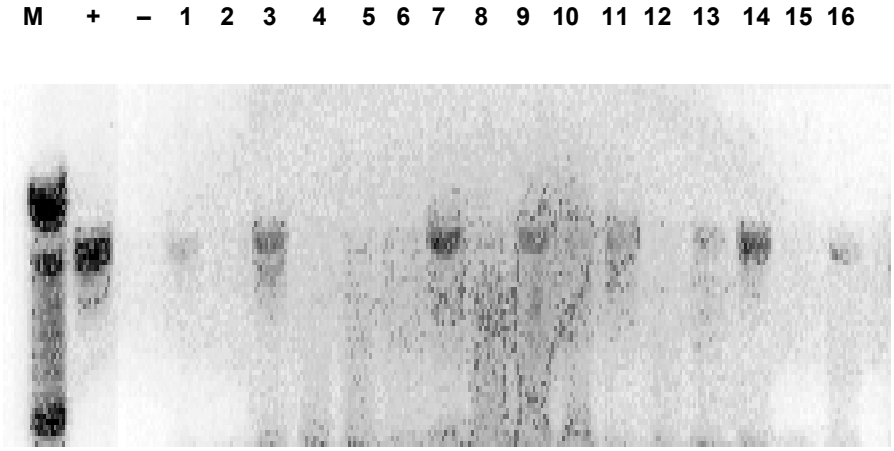


Figure 3.5 Long term propagation of TtARS1 in *Tetrahymena*.

PCR analysis with pCC1FOS vector specific primers was done on total genomic DNA extracted from *Tetrahymena* transformants of TtARS1 propagated to 85 fissions. M represents 100 bp marker, + is plasmid positive control for PCR, - is negative control, and 1-16 represents PCR products from 16 transformants of TtARS1. 9 out of 16 transformants produced PCR product and hence, they propagate TtARS1 until 85+ fissions.

the rest of the transformed cells were propagated as an en masse culture. Genomic DNA was extracted from 50 ml cultures of *Tetrahymena* en masse transformants at 15, 30, 45 and 60 fissions. The DNA was digested with BamHI, since there are two BamHI sites in the pCC1FOS vector flanking the insert and no BamHI sites in TtARS1 insert. The upper left panel of figure 3.6 shows a graphic representation of BamHI digestion products derived from the endogenous TtARS1 chromosomal fragment and episomal TtARS1. Southern blot (lower panel of figure 3.6) was done using the respective insert fragments as radiolabelled probes to enable the comparison of the abundance of the episomal copy relative to the endogenous chromosomal segment. The Southern blotting results showed two bands, a 15 kb chromosomal band and an episomal band (4 kb for ΔNhe-1 and 5 kb for ΔNsi-1). This indicated that the 5.0 kb ΔNsi-1 and the 4.2 kb ΔNhe-1 fragments of TtARS1 are capable of autonomous DNA replication in *Tetrahymena* for at least 60 fissions. The copy number of the ΔNhe-1 and ΔNsi-1 episome was similar to that of the endogenous chromosome as was seen before for the full length TtARS1 fragment. Since, the internal deletions in TtARS1 did not compromise the functioning of the replicon, all the cis-acting determinants for origin function are contained within the 4.2 kb ΔNhe-1 fragment (note: the ΔNhe-1 deletion construct has been emphasized here since the deleted fragment in ΔNhe-1 is larger than ΔNsi-1).

Transfection frequency

A portion of the transformed culture was plated at different dilutions in 96 well microtiter dishes. The plates were scored to calculate the total number of paromomycin

Deletion construct	Type of digest	Episomal fragment	Chromosomal fragment
ΔNheI	BamHI	4.0 kb	15.0 kb
ΔNsiI	BamHI	5.0 kb	15.0 kb

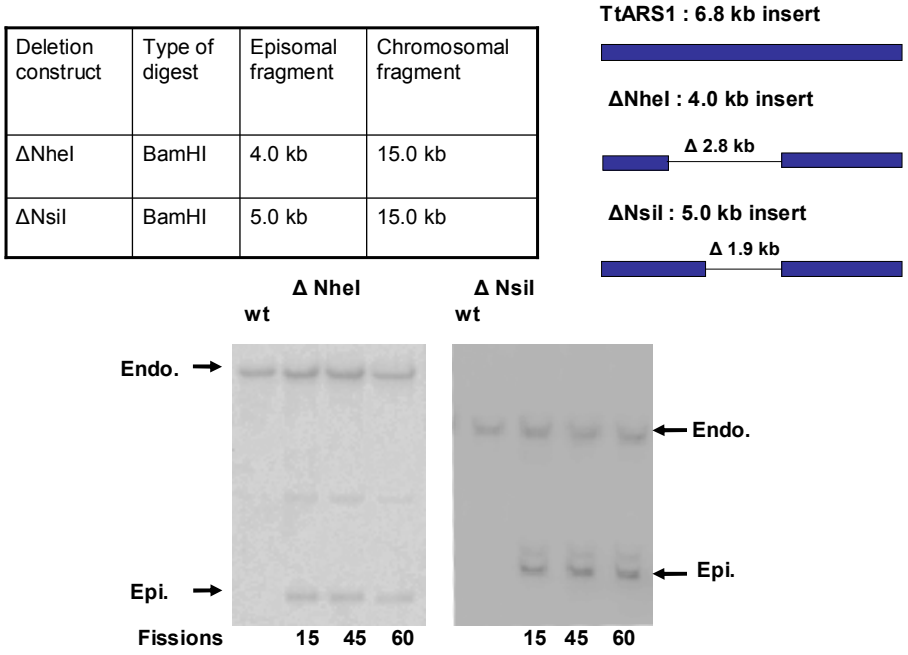


Figure 3.6 The deletion derivatives of TtARS1, Δ Nhe1 and Δ Nsi1 are maintained in *Tetrahymena* as episomes.

Table on top left indicates the different size fragments produced after BamHI digestion of genomic DNA from Δ Nhe1 and Δ Nsi1 *Tetrahymena* transformants. Panel on the right shows the size of internally deleted fragments in the deletion constructs. Southern blot analysis was done using total genomic DNA from the respective transformants at the mentioned fissions. The respective *Tetrahymena* insert fragments were used as radiolabelled probes. Endo. represents the signal from endogenous native chromosome and epi. represents the signal from the episome. wt represents DNA from untransformed *Tetrahymena* cells which show the presence of endogenous signal only. The calculated transformation frequencies were 3920 transformants/μg of transforming DNA with ΔNhe-1, and 3680 transformants/μg of DNA in case of ΔNsi-1.

resistant wells and the transfection frequency was calculated as the total number of transformants obtained per microgram of transforming DNA. The transformation frequency with ΔNhe-1 transformation was 3920 transformants/μg of transforming DNA, and in case of ΔNsi-1, it was 3680 transformants/μg of DNA.

Plasmid re-isolation assay: Plasmid DNA was extracted from both, ΔNhe-1 and ΔNsi-1 en masse transformants at passages 2 and 4. The extracted plasmid was digested with Dpn*I* to remove any residual input DNA and re-transformed into *E. coli*. This approach generated *E. coli* transformants, which reconfirmed that Nsi-1 and Nhe-1 can be propagated as episomes.

Internal fragments are not capable of supporting autonomous DNA replication

The next aim was to determine whether the internal fragments of 1.9 kb (ΔNsi-1) and 2.8 kb (ΔNhe-1) fragments support autonomous replication in *Tetrahymena*. These internal fragments were blunt end cloned into the pSMART vector and the newly generated constructs were designated ΔNsi-2 and ΔNhe-2 respectively. The maintenance of Nsi-2 and Nhe-2 was studied by transfecting these plasmids into *Tetrahymena* with AN101 as the co-transforming vector. The protocol for cell starvation, transfection and re-feeding of transformants was as previously described. Transformants were selected for paromomycin resistance and a portion of the cells were plated into 96 well dishes. The rest was cultured and passaged en masse until 25 fissions.

Southern blot analysis of en masse culture: The purpose of the en masse culture was to study the maintenance of the deletion derivatives and compare the copy number with the endogenous native chromosomal copy. Genomic DNA was extracted from en

masse cultures of the Nsi-2 and Nhe-2 transformants at 15, 22 and 30 fissions. 50 μg of the DNA was digested with EcoRV enzyme. The upper panel of figure 3.7 shows a graphic representation of EcoRV digestion products derived from the endogenous chromosomal fragment and the plasmid fragment. The digested DNA was run on a 0.7% agarose gel and transferred to a Genescreen plus membrane. Southern blot hybridization was performed using the respective insert fragments as a probe (lower right panel of Figure 3.7). Southern blotting with the insert fragment probe produced a single band that corresponds to the chromosomal in *Tetrahymena* transformants of Nhe-2 and Nsi-2. Simultaneously, the pSMART vector fragment was used as probe for confirmation (lower left panel of Figure 3.7). There was no signal in the lanes that contained genomic DNA from the transformants of deletion derivatives. The lane containing genomic DNA from *Tetrahymena* transformants of TtARS1 showed a single band, which indicated once again that TtARS1 is capable of autonomous DNA replication (Figure 3.7). The absence of any signal relating to the episome indicates that the internally deleted fragments were unable to support autonomous DNA replication.

Transfection frequencies were calculated as an indicator of successful transformation of DNA into cells. The efficiency of transfection was comparable to the prior experiments, with 3246, 3360 and 4080 transformants/μg of transformed DNA for TtARS1 control, Nhe-2 and Nsi-2, respectively. Thus, the above experiment showed that two overlapping internal fragments of 2.8 kb and 1.9 kb did not support autonomous replication. Since the internal deletions within TtARS1 did not affect replication efficiency, hence at this point the preliminary conclusion was that the cis-acting

Deletion construct	Type of digest	Episomal fragment	Chromosomal fragment
Nhe2	EcoRV	2.8 kb	9.0 kb
Nsi2	EcoRV	1.9 kb	9.0 kb

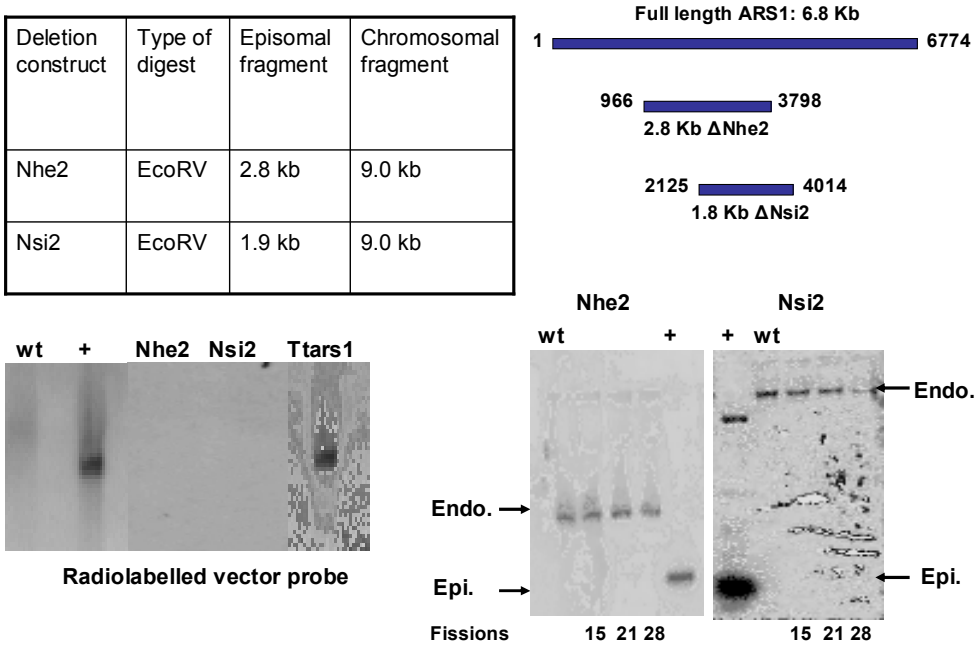


Figure 3.7 Internal fragments of TtARS1 do not support autonomous DNA replication.

Table on upper left panel indicates the different size fragments produced after EcoRV digestion of genomic DNA from Nhe-2 and Nsi-2 *Tetrahymena* transformants. Upper right panel shows the insert size of the deletion constructs Nhe-2 and Nsi-2. Panel on bottom left indicates Southern blot data using total genomic DNA from the transformants and with the respective vector fragments as radiolabelled probe. wt represents genomic DNA from wild type untransformed *Tetrahymena* cells, + represents plasmid positive control, Nhe-2 and Nsi-2 represent genomic DNA from *Tetrahymena* transformants of Nhe-2 and Nsi-2 respectively, and TtARS1 is genomic DNA from full length TtARS1 transformants of *Tetrahymena*. Bottom right panel shows Southern blot data using the respective insert fragments as radiolabelled probe. Endo. represents endogenous chromosomal signal and epi. represents signal from the plasmid. + is plasmid positive control and wt is genomic DNA from untransformed *Tetrahymena* cells. Transfection frequencies calculated for this experiment were 3246, 3360 and 4080 transformants/µg of transformed DNA for TtARS1 control, Nhe-2 and Nsi-2, respectively

determinants required for origin function was either in the left piece or the right piece in context to the Nhe-2 internal fragment.

The 3.0 kb Nhe-R1 fragment supports autonomous DNA replication

The next step was to determine whether the origin is in the left or the right fragment with respect to the deletion (2.8 kb internal deletion in ΔNhe-1). The 966 bp left fragment was designated Nhe-L. This fragment was in the pCC1FOS vector backbone (8.1 kb) and hence the total size of Nhe-L construct was 9.0 kb. The 3.0 kb right fragment was designated Nhe-R1. This fragment was cloned into pSMART vector. The first derivative to be studied was Nhe-R1, which was co-transfected into the developing macronucleus of *Tetrahymena* with AN101 plasmid. TtARS1 was co-transfected with AN101 as positive control. Transformants were selected with paromomycin (100 μg/ml) and divided into two portions as before. One portion was plated at different dilutions (undiluted, 1:10 and 1:100) in 96 well microtiter dishes. The second portion of transformants was cultured en masse for Southern blot analysis. Genomic DNA was extracted from the en masse propagated cultures at fission 15, 30, 45 and 56 and digested with EcoRI since there are two EcoRI sites flanking the insert in the pSMART vector. Figure 3.8 provides a graphical representation of fragments produced after EcoRI digest of genomic DNA from Nhe-R1 transformants. EcoRI digestion produces a 9.0 kb chromosomal fragment and a 3.0 kb episomal fragment (left panel of Figure 3.8). The fragments were separated in a 0.7 % agarose gel, transferred to a nylon membrane and hybridized with Nhe-R1 insert fragment as probe. The analysis revealed that the Nhe-R1 construct was maintained as an episome for 60 fissions in *Tetrahymena*

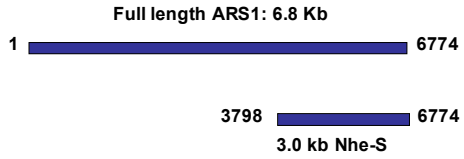
(right panel of Figure 3.8). Genomic DNA from wild type untransformed cells showed the presence of only the chromosomal band and no episomal band. Thus, the 3.0 kb Nhe-R1 deletion derivative contains cis-acting sequences that are sufficient to support stable autonomous replication in *Tetrahymena*.

As a second approach to confirm these results, plasmid DNA was extracted from Nhe-R *Tetrahymena* transformants at fission 15 and 30, digested with Dpn*I* and re-transformed into *E. coli*. This approach yielded *E. coli* transformants which confirmed that Nhe-R1 deletion derivative replicated in *Tetrahymena* and could be retrieved back a circular episome. The transfection frequency for this experiment was calculated to be 4880 transformants/μg of transforming DNA. Thus, from the above experiments it can be concluded that TtARS1 contains at least one non-rDNA replicon, which has been delineated to a 3.0 kb fragment. The next step was to further delineate the functional replicon within this 3.0 kb Nhe-R fragment.

A 1.2 kb region within Nhe-R1 contains the replicon

Most eukaryotic replicons that have been identified so far are located at intergenic or non-coding sequences of DNA. Recently, a gene prediction database for *Tetrahymena* genome has been developed. The analysis of TtARS1 region predicted two coding sequences and two intergenic regions (Figure 3.9). One of the intergenic regions is 1.2 kb in length and is contained within the 3.0 kb Nhe-R fragment that functions as a replicon. The second intergenic region is 599 bp in length and 300 bp of this region is contained within TtARS1. The rest of the 599 bp intergenic region extends into the native chromosome and is not included in the episomal TtARS1. Next, I chose to assess

Deletion construct	Type of digest	Episomal fragment	Chromosomal fragment
Nhe-R1	EcoRI	3.0 kb	9.0 kb



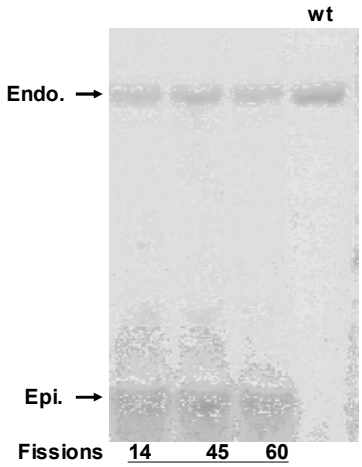


Figure 3. 8 A 3.0 kb deletion derivative of TtARS1 called Nhe-R1 can replicate autonomously in *Tetrahymena*.

Top left panel shows the different size fragments produced after EcoRI digestion of *Tetrahymena* transformants of Nhe-R. The sizes of insert in Nhe-R1 as compared to full length TtARS1 is shown in the bottom left panel. Panel on the right is Southern blot data using total genomic DNA from Nhe-R transformants at the mentioned fissions. Nhe-R1 insert fragment was used as radiolabelled probe. Endo. represents the signal from endogenous native chromosome and epi. represents the signal from the episome. wt represents DNA from untransformed *Tetrahymena* cells. The calculated transfection frequency for this experiment was 4880 transformants/µg of transforming DNA for Nhe-R1 and 3996 transformants/µg of transforming DNA for TtARS1 control.

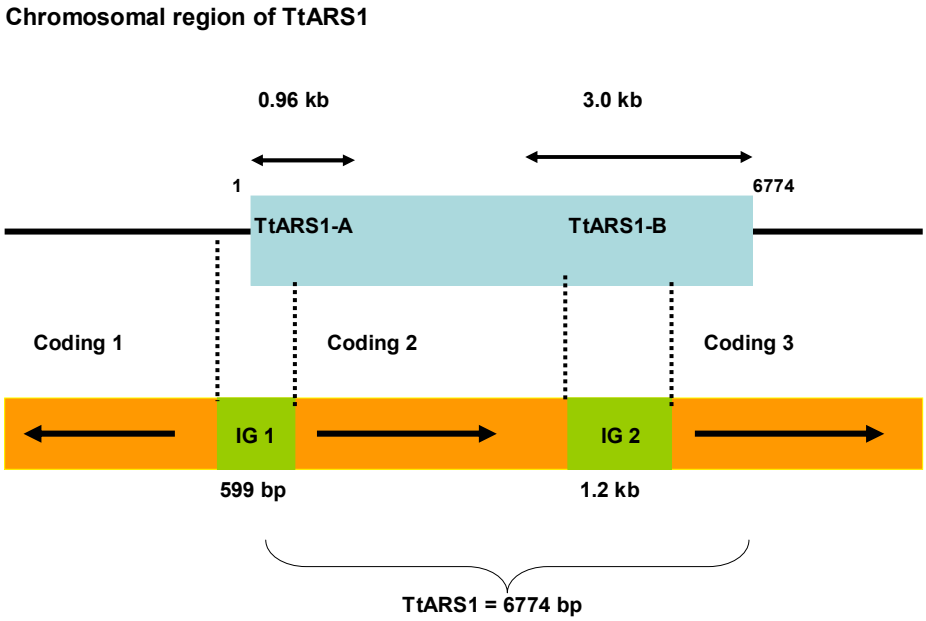


Figure 3.9 Gene prediction analysis of TtARS1 chromosomal segment.

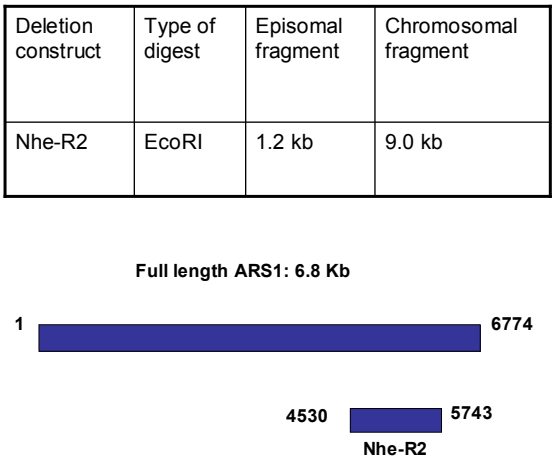
Green regions depict predicted intergenic / non-coding sequences, while the orange regions depict coding sequences. There are 2 intergenic regions, one in each functional segment of TtARS1. IG1 is partially contained (300 bp) in TtARS1-A, and IG2 is completely present within TtARS1-B. There are 2 predicted coding regions within TtARS1. The arrows denote the direction of transcription.

whether the 1.2 kb segment within the 3.0 kb Nhe-R1 fragment is sufficient to support autonomous replication. This segment was amplified by PCR and cloned into pSMART vector, and the construct was designated Nhe-R2. *Tetrahymena* was co-transformed with Nhe-R2 and AN101 and paromomycin-resistant transformants were propagated en masse until 60 fissions. Genomic DNA was extracted from *Tetrahymena* transformants at fission 15, 30, 45 and 60, digested with EcoRI and Southern blot was carried out using the 1.2 kb insert fragment as radiolabelled probe. EcoRI digestion produces a 9.0 kb chromosomal fragment and a 1.2 kb episomal fragment (left panel of Figure 3.10). Southern blotting results showed the presence of the episomal fragment in addition to the chromosomal fragment in *Tetrahymena* transformants of Nhe-R2 (right panel of Figure 3.10). This indicates that Nhe-R2 is capable of autonomous replication as an episome in *Tetrahymena* for 60 fissions and hence all the cis-acting determinants of the replicon are contained within the 1.2 kb Nhe-R2 intergenic region.

The following is a summary of my results thus far:

1) A 6.8 kb fragment of *Tetrahymena* genome was isolated (TtARS1) that supports autonomous replication.
2) Derivatives of TtARS1 that contains internal deletions of 1.9 kb and 2.8 kb (ΔNhe-1 and ΔNsi-1, respectively) also support autonomous replication.
3) The internal fragments. Nhe-2 and Nsi-2 are not capable of maintenance in *Tetrahymena*.

Deletion construct	Type of digest	Episomal fragment	Chromosomal fragment
Nhe-R2	EcoRI	1.2 kb	9.0 kb



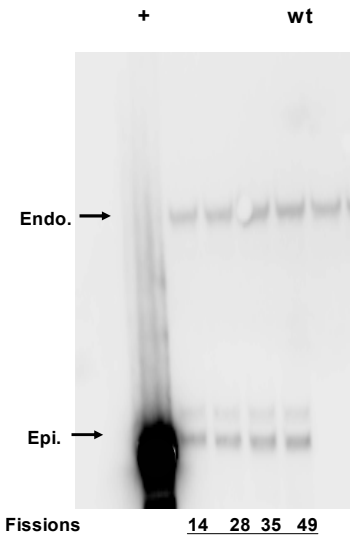


Figure 3.10 A 1.2 kb segment of predicted intergenic region within Nhe-R2 functions as an autonomous replicating sequence in *Tetrahymena*.

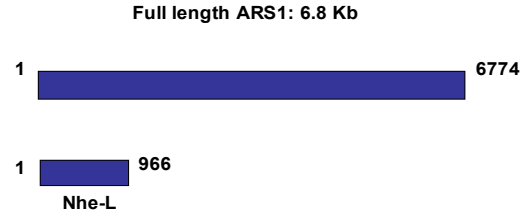
Table in top left panel shows the different size fragments produced from the episomal and chromosomal loci after EcoRI digest of total genomic DNA from *Tetrahymena* transformants of Nhe-R2. Bottom left panel indicates the size of Nhe-R2 in comparison to full length TtARS1. Panel on the right is the Southern blot data using EcoRI digested genomic DNA from Nhe-R2 transformants of *Tetrahymena*. Endo. and Epi. represent signals from endogenous chromosomal locus and episomal locus respectively. + indicates plasmid positive control and wt is genomic DNA from untransformed cells. Transfection frequencies for this experiment was 4996 transformants/µg of transforming DNA for Nhe-R2 and 4110 transformants/µg of transforming DNA for TtARS1 control.

4) The deletion construct containing 3.0 kb fragment of TtARS1 (Nhe-R1) replicates in *Tetrahymena*. Further deletion of Nhe-R1 revealed that a 1.2 kb intergenic fragment designated Nhe-R2 also supports autonomous replication. The next step was to determine whether the left 966 bp sequence of TtARS1 with respect to the internal deletion (ΔNhe-1 construct) has replicator function.

TtARS1 contains 2 independent replicons

To test whether the left end of the internally deleted derivatives contain a second replicon, I tested the replication competence of the 966 bp fragment, designated Nhe-L. *Tetrahymena* cells were co-transformed with AN101 and Nhe-L. The protocol for transfection, determination of replication competence and the transfection frequency remained the same as the above experiments. TtARS1 was employed as positive control for transfection. Total genomic DNA was extracted from the co-transformants at fissions 15, 30, 45 and 60, digested with BamHI. Figure 3.11 shows the different fragments produced after BamHI digest of genomic DNA from Nhe-L transformants. The probe Nhe-L insert fragment) hybridized to two bands, one corresponding to the endogenous chromosome and the other to the episomal copy. The copy number of the episome was similar to that of the endogenous chromosome. The result indicates that this 966 bp fragment contained cis-acting sequences that support autonomous DNA replication. Thus, I conclude that TtARS1 has two independent replicons, one on either side of the 2.8 kb internal deletion by NheI enzyme. The left replicon is contained within a 966 bp fragment and the right replicon is within a 1.2 kb fragment. The two replicons were designated as TtARS1-A (Nhe-L) and TtARS1-B (Nhe-R2).

Deletion construct	Type of digest	Episomal fragment	Chromosomal fragment
Nhe-L	BamHI	966 bp	15.0 kb



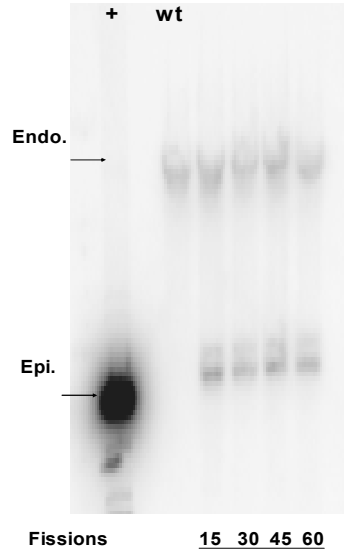


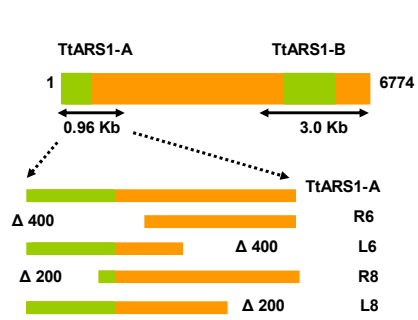
Figure 3. 11 TtARS1 contains two independent replicons.

The deletion derivative Nhe-L, which contains 966 bp of TtARS1 inserts, is maintained in *Tetrahymena* as an episome. Panel on top left shows the different size fragments produced after BamHI digestion of *Tetrahymena* transformants of Nhe-L. Right panel shows Southern blot analysis using total genomic DNA from Nhe-L transformants at the mentioned fissions. Nhe-L insert fragment was used as radiolabelled probe. Endo. represents the signal from endogenous native chromosome and epi. represents the signal from the episome. wt represents DNA from untransformed *Tetrahymena* cells and + indicates BamHI digested plasmid as positive control. Transfection frequencies were 3850 transformants/µg of DNA for Nhe-L and 3992 transformants/µg of DNA for TtARS1 control.

The next aim was to characterize the origin and identify the minimal size and the cis-acting determinants that are necessary for replicon function. Nhe-L fragment was chosen for further dissection because of its small size. The classical approach of deletion analysis was followed for this characterization.

Deletion analysis of TtARS1-A replicon

TtARS1-A resides within a 966 bp *Tetrahymena* fragment in the 8.1 kb pCC1FOS vector. To determine the minimal replicon size, deletion constructs L6, R6, C6, L8 and R8 were generated. These derivatives are represented in Figure 3.12. These derivatives were separately co-transfected into the developing macronucleus of mating cells of *Tetrahymena* with the plasmid AN101. Total genomic DNA from en masse propagated paromomycin-resistant transformants was extracted at fissions 15, 30, 45 and 60 and analyzed for episomal DNA replication. Southern blotting (Figure 3.12) revealed that the L8 construct, which contained 800 bp sequence from the left end of TtARS1-A insert, replicated autonomously in *Tetrahymena* and was stably maintained at a level similar to the native chromosome. In contrast, the deletion construct R8 failed to support autonomous replication. Comparison of L8 and R8 constructs indicate that the leftmost 200 bp fragment that is present only in L8 may contain one or more essential determinants. This 200 bp fragment lies within a predicted intergenic region. Southern blot analysis also revealed that 400 bp deletions from either side of TtARS1-A abolished origin activity, as shown for the deletion constructs R6 and L6. As expected, the central 600 bp fragment (C6) of TtARS1-A failed to support autonomous replication (data not



Replication competence

Constructs	Insert size	Replication competence
TtARS1	966 bp	Competent
R6	600 bp	Incompetent
R8	800 bp	Incompetent
L6	600 bp	Incompetent
L8	800 bp	Competent

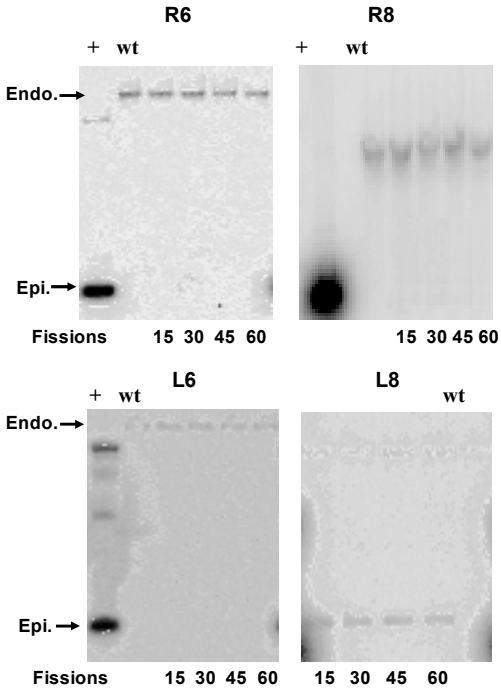
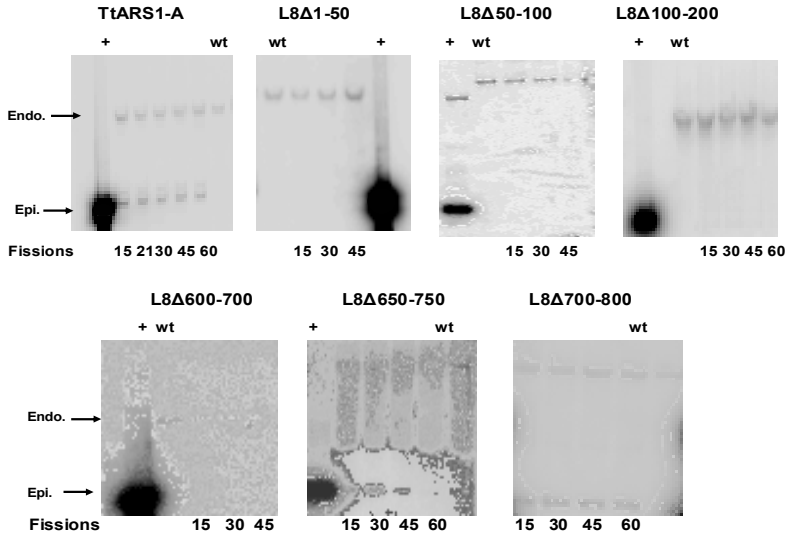


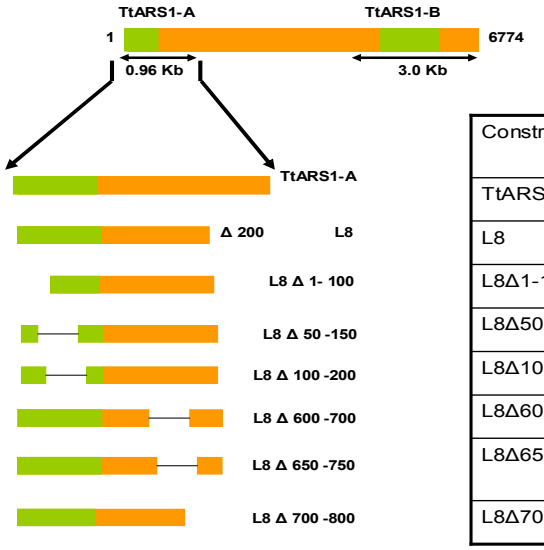
Figure 3.12 Origin is contained within 800 bp segment of TtARS1-A.

Top left panel shows the deletion derivatives of TtARS1-A, which functions as an episome in *Tetrahymena*. The green region signifies predicted non-coding or intergenic sequences and the orange region signifies coding sequences. Right panel shows Southern blot data using BamHI digested, total genomic DNA from *Tetrahymena* transformants of the respective constructs and with the respective insert fragments as radiolabelled probe. Endo. indicates signal from native chromosome and epi. indicates signal from episome. Table in bottom left panel shows a summary of the Southern blot data, stating the replication competency of the deletion derivatives. Transfection efficiencies were 4440, 3964, 4903, 3375 and 3872 transformants/μg of transforming DNA for TtARS1, R6, R8, L6 and L8 respectively.

shown). In summary, the two 200 bp segments at the left and right ends of L8 are necessary for autonomous replication. The fact that L8 construct replicates in *Tetrahymena* and L6 does not leads to two speculations. The first one is that specific cis-acting determinants are present in the 200 bp segment that has been deleted in L6. This would be interesting since it suggests for the first time that coding regions have information that contributes to origin function. The second is that the replicon function may be constrained by the decreased size of the insert, which results in a reduced AT content. To distinguish between the two possibilities and to further delineate the TtARS1-A replicon, a set of deletion derivatives of L8 was generated that contained 100 bp deletions, overlapping one another by 50 bp. A graphical description of the constructs is in Figure 3.13. The deletion constructs were designated as L8Δ1-100, L8Δ50-150, L8Δ100-200, L8Δ600-700, L8Δ650-750 and L8Δ700-800. They were co-transformed into the developing macronucleus of *Tetrahymena* with AN101 plasmid. Southern blot analysis of genomic DNA at fissions 15, 30, 45 and 60 revealed that four of the constructs, L8Δ1-100, L8Δ50-150, L8Δ100-200 and L8Δ600-700 were not replicated in *Tetrahymena* (Figure 3.13). No hybridization signal was detected even at the earliest time point. In contrast, TtARS1 and L8Δ600-700 were stably maintained in *Tetrahymena*.

The deletion derivative, L8Δ650-750 gave a very interesting result. In the initial *Tetrahymena* passages (fission # 15 and 30), L8Δ650-750 was able to support autonomous replication. But in the later passages (fission 45+), the plasmid was





Constructs	Insert size	Replication competence
TtARS1-A	966 bp	Competent
L8	800 bp	Competent
L8Δ1-100	700 bp	Incompetent
L8Δ50-150	700 bp	Incompetent
L8Δ100-200	700 bp	Incompetent
L8Δ600-700	700 bp	Incompetent
L8Δ650-750	700 bp	Partially competent
L8Δ700-800	700 bp	Competent

Figure 3.13 Origin is mapped to a 700 bp segment within L8.

Top panel indicates Southern blot data using BamHI digested genomic DNA from *Tetrahymena* transformants of the respective constructs. The respective inserts were used as radiolabelled probes. Endo. and Epi. represent signal from endogenous chromosome and episome respectively. + is plasmid positive control and wt is genomic DNA from untransformed *Tetrahymena* cells. Bottom left panel shows the different deletion derivatives of L8. Each of these has overlapping deletions of 100 bp within two 200 bp segments that has been shown to be essential for replicon function. Bottom right panel provides a summary of the size and replication competence of the different deletion derivatives. Transfection efficiencies were comparable to prior experiments.

gradually lost from the transformants. This strongly suggests the presence of cis-acting determinants in TtARS1-A that regulate origin function. The fact that all the deletion derivatives are the same size, some of which replicate and others that do not, rules out the possibility that replication competence is simply a function of insert size.

Analysis of intergenic regions from TtMac library

The starting material for identifying non-rDNA origins of replication in *Tetrahymena* was a macronuclear sub-library that represented 0.6X coverage of the *Tetrahymena* genome. From this sub-library, plasmid DNA was prepared from a pool of 100 colonies that lacked the rDNA 5' NTS tested for origin function in *Tetrahymena*. The reported percentage of empty vector in the library was 5%. Considering that 95% of the library had inserts of 5–7 kb length, 100 library clones would represent approximately 500–700 kb of DNA. In spite of analysis of such an expansive DNA stretch, only one 6.8 kb segment, TtARS1, was identified that supported autonomous replication. This low rate of isolation of a functional DNA segment leads to many speculations. One possibility is that the 5–7 kb inserts are too small to encompass most replicons completely. Therefore, only parts of replicons may have got cloned, leaving out necessary functional determinants. Another possibility is that TtARS1 replicated more robustly than other replicons in the plasmid DNA pool. Hence, in the initial ARS isolation assay, where plasmid from *Tetrahymena* was re-transformed into *E. coli*, only TtARS1 was isolated. One way to address this issue is to test the ability of different intergenic regions in the TtMac library to support autonomous replication.

The number of empty vectors in the TtMac library was reported to be below 5% by the manufacturer. This was re-tested by plating cells derived from glycerol stocks of sub-library # 29, which was used to isolate TtARS1, onto LB + chloramphenicol plate and isolating plasmid DNA from individual colonies. Twenty-eight representative clones were sequenced using vector primers that flank the insertion site. Sequencing of the clones revealed that only ten out of twenty eight clones tested contained inserts. The remainder contained empty vectors. Thus the average number of clones in the library that contained *Tetrahymena* inserts was 35%. Gene prediction of the ten sequenced inserts revealed that only Δfive contained predicted intergenic regions. The respective sizes of the intergenic regions are listed in Table 2.1 of methods chapter. All these inserts contained sequences that were different from TtARS1 and rDNA sequence.

Plasmid DNA from these five library clones was pooled to transfect into the developing macronucleus of *Tetrahymena*. The transformants were selected with paromomycin (100μg/ml) since AN101 plasmid was used as the co-transforming vector. TtARS1 was transfected in *Tetrahymena* as a positive control for transfection. The transformants were propagated en masse for 15 and 30 fissions of *Tetrahymena*. Total genomic DNA was extracted from the transformants for Southern blot analysis using the pCC1FOS vector fragment as radiolabelled probe (Figure 3.14). Uncut DNA was run on a 0.7% agarose gel for this purpose. Southern blotting results showed no evidence for episomal DNA in the five clone pool. TtARS1, which was used as positive control, was maintained as an episome in *Tetrahymena*. This result implies that the inserts in the TtMac clones probably did not contain cis-acting elements that are necessary for origin

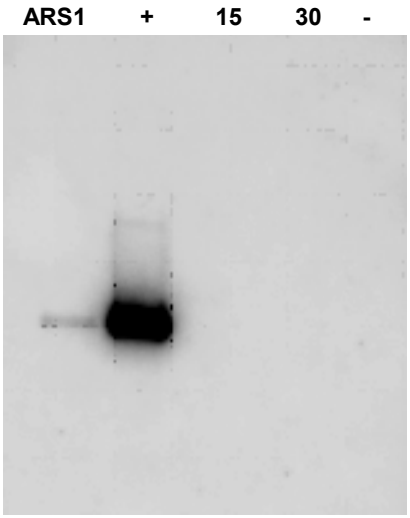


Figure 3.14 Pool of newly isolated predicted intergenic regions in TtMac library fail to support autonomous DNA replication.

Southern blot analysis with total uncut genomic DNA (extracted from fission # 15) was done using pCC1FOS vector fragment as radiolabelled probe. ARS1 indicates DNA from *Tetrahymena* transformants of TtARS1. + is plasmid positive control. – indicates negative control for transfection (genomic DNA from *Tetrahymena* transformants of Nhe2 deletion construct). 15 and 30 represents genomic DNA from 15th and 30th fission of *Tetrahymena* transformants containing TtMac library plasmid (pool of 5 clones) with intergenic regions. Transfection efficiencies were 4620 transformants/µg of DNA for TtARS1 control and 4998 transformants/µg of DNA for the TtMac library plasmid pool.

function. However, the intensity of the signal for TtARS1 transformants was faint (Figure 3.14). If the transforming plasmid pool contained one clone that replicated in *Tetrahymena*, one might not detect the signal, since the intensity would be one-fifth of the TtARS1 signal (Figure 3.14). Moreover, if replication was not as competent as TtARS1, then the signal would be almost negligible. Hence, these results are not definite and it is necessary to study each intergenic region in the transforming DNA pool separately by transfecting *Tetrahymena*. Also, a more thorough analysis of the TtMac library is required to isolate new replication origins and preferably a *Tetrahymena* library must be used that contain larger size inserts.

Bioinformatic analysis of TtARS1

TtARS1 has been shown to contain two independent replicons, designated TtARS1-A with a minimal size of 700 bp and TtARS1-B, which is characterized to a size of 1.2 kb. In an effort to identify conserved sequences that might be binding sites for trans-acting factors, I performed a bioinformatic analysis of TtARS1 and rDNA replicons. The rDNA replicon contains discrete cis-acting sequences that have been shown to be essential for initiation of replication. As a first step for the bioinformatic analysis, a search for these cis-acting elements (namely type I elements and parts of the pause site elements) was performed in TtARS1 by using ClustalW program. This analysis showed no significant similarity between the rDNA and TtARS1 replicons. LALIGN program was used to identify similar sequences between rDNA origin (5' NTS) and TtARS1-A and TtARS1-B. Table 3.2 provides information on the sequence identity between the different queries. This analysis was done in both the forward and

reverse complementary orientations of the sequences (Table 3.2). The results showed that there are identical stretches of six to nine basepairs between rDNA origin and TtARS1. This was true for both orientations. Next, analysis of TtARS1-A and TtARS1-B was done to find identical sequences between them (Table 3.2). As before, there were sequences of eight to nine basepairs that were identical between the two non-rDNA replicons, irrespective of the orientation of the inserts. Genetic dissection of TtARS1-A has indicated that two segments of 200 bp at the left and right end of L8 construct are necessary to support autonomous replication. LALIGN was used to identify similar sequences between these two 200 bp sequences (Table 3.2). Three identical DNA sequences between seven to nine basepairs were identified. There were no identical stretch of sequences between the 200 bp segment and the rDNA 5' NTS. As a control, four different arbitrary sequences of *Tetrahymena* were used as queries for the LALIGN program. These segments represented coding and non-coding sequences of *Tetrahymena*, as determined by gene prediction. Data is shown for only one pair of non-coding sequences (Table 3.2). Identical sequences of eight to ten basepairs were detected in these arbitrary sequences. However, this result is not surprising, considering the high AT content and repetitive AT sequences in *Tetrahymena.* Thus, the bioinformatic analysis of TtARS1 did not provide any definitive result with respect to sequence conservation.

The frequencies of bases in the genome of an organism are not always equiprobable. A particular stretch of DNA can have high AT content relative to GC content. The skew is calculated as (nucleotide1 - nucleotide2) / (nucleotide1 +

Table 3. 2 Bioinformatic analysis of non-rDNA and rDNA replicons.

The program, LALIGN was used to look for sequence conservation between rDNA and non-rDNA replicons. RC indicates the reverse complement of the indicated DNA sequence. For example, in the second column, TtARS1-A in the forward orientation was used as a query against the reverse complement of TtARS1-B. The two arbitrary *Tetrahymena* sequences were of 1.0 kb length and represented non-coding DNA sequence.

QUERY	SEQUENCE IDENTITY
TtARS1-A and rDNA 5'NTS	GTATTTAAA
	ATCAAAA
TtARS1-A and Rdna 5'NTS – RC	AATTAAATT
	TTTAAAAT
	TTTAAAAT
	CAAATTT
	GCTTTTT
	ATTTTTT
	TTTAACT
TtARS1-B and rDNA 5'NTS	TTTTTTTTA
	AAAAAAT
	ATTTTTTA
	AAATTTA
TtARS1-B and rDNA 5'NTS – RC	GATTTTTT
	AAAATTT
	AAATATT
	TTTTTGT
TtARS1-A and TtARS1-B	ATTTAAAAAA
	ATTTAAAAA
	GTCTAATTT
	TTTTTAGAT
	GTATTTAA
	GCAAAAA
TtARS1-A and TtARS1-B – RC	AATATTTTA
	AAATTAGA
	GTTAATAA
	ATTTAAAA
	CTAAAAA
	AATTGCA
L1-200 and L600-800	TTTTTAGAT
L1-200 and L600-800 – RC	TTTTCAATT
	TAACAAT
Two arbitrary non-coding *Tetrahymena* sequences	ATATTTTAAA
	TTTTTTAAATT
	TATTTCAAA
	TAATTATAA
	TTCTTTTTT
	TAATTATT

nucleotide2). In *S. pombe*, AT-skew analysis has been used to predict potential sites in the chromosomes that can function as origins (19). All *S. pombe* origins that have been identified thus far are located in intergenic regions. The AT contents of intergenes are much greater than those of genes. It has been observed in *S. pombe* intergenic regions, there is an AT skew such that the probability that an A residue will be followed by another A residue is significantly greater than expected on a random basis (19). Significantly, the same general pattern was observed for *S. pombe* ARS elements. Thus, the sequence properties of ARS elements are quite similar to those of intergenes in general. Also, the ARS elements show base asymmetry and are enriched in runs of A's and T's. Thus, it has been speculated that a high proportion of *S. pombe* intergenes might be capable of functioning as origins of DNA replication (19). A similar study was done with TtARS1 for fine mapping of the non-rDNA replicons. Also, AT-skewness of the *Tetrahymena* genome was studied to look for putative origins. However, due to the high AT content of *Tetrahymena* genome, the AT-skew did not vary largely between predicted intergenic regions and coding sequences. As a result of this, within the scope of our study, it was not feasible to use AT-skewness as a criterion to look for replicons.

As a summary of this study, a non-rDNA replicon, designated TtARS1, was identified in *Tetrahymena*. Genetic dissection of the replicon revealed two independent origins of replication, TtARS1-A of less than or equal to 700 bp and TtARS1-B of 1.2 kb, which are sufficient to support autonomous replication and long term maintenance of episomal DNA. Bioinformatic analysis of TtARS1 revealed no significant similarity between the two non-rDNA replicons, or between rDNA and either non-rDNA replicon.

CHAPTER IV

SUMMARY AND DISCUSSION

Chromosome replication is a key event in the cell division cycle in all organisms. Replication initiates at multiple origins throughout S phase in a highly coordinated manner, such that no region of the genome is left unreplicated and no region is replicated more than once. Studies over the past few years have led to the identification of several eukaryotic replicons, and much of the work has been done in the genetically tractable *S. cerevisiae*. There are a handful of metazoan replicons that have been identified, but due to their high complexity very few have been genetically dissected and functionally characterized (151). Moreover, it has been observed that although the trans-acting factors and basic biochemical mechanism of replication initiation in eukaryotes are conserved, the DNA sequences that promote these events are unique in different organisms. Cis-acting elements are much dispersed in metazoan replicons, ranging from hundreds to thousands of basepairs (1). In addition, there are epigenetic factors such as chromatin modification that play a role in selecting sites in chromosomes for origin function (118, 119). Comparative studies between the limited numbers of metazoan replicons that have been characterized thus far have failed to identify any consensus sequences that direct the targeting of replication specific proteins (152). It is worth noting that *S. pombe* and metazoan ORCs bind DNA non-specifically, preferentially interacting with blocks of AT-rich degenerate sequences (15,17). In contrast, *S. cerevisiae* replicons as well as the genetically characterized *Tetrahymena* rDNA replicon

contain discrete cis-acting determinants, such that point mutations can affect replication function (59, 63, 153). Thus, it is evident that the specification of replication origins varies to a great extent in different organisms and this denotes the difficulty in the identification and study of replicons in higher eukaryotes.

The identification of replicons in *S. cerevisiae* was through the isolation of DNA segments that support autonomous replication of episomes (ARS assays) (154). In higher eukaryotes this is challenging since DNA segments that function as autonomous replicons are often not functional in the native chromosome. As an example, in humans, it has been shown that any DNA sequence that is >15 kb can replicate autonomously at ectopic loci (155). The rDNA minichromosome in *Tetrahymena thermophila* has served as a useful model for studying complex origins of DNA replication. The rDNA minichromosome is specialized in that it undergoes locus specific gene amplification. This minichromosome exists as a 21 kb palindrome and each origin is localized in a 1.9 kb region of the rDNA (52, 53), which consists of several discrete cis-acting elements that can confer autonomous replication to plasmids. In spite of an extensive knowledge of the replicon structure, the mechanism of selective gene amplification remains to be elucidated. The lack of knowledge of other replicons in *Tetrahymena* makes comparative analysis impossible.

Isolation and functional analysis of a non-rDNA replicon in *Tetrahymena*

The goal of my thesis research was to develop a method for identifying new replication origins in *Tetrahymena*. To that effect, I have developed a DNA transformation assay, designated as the ARS isolation strategy, to isolate new replication

origins in *Tetrahymena*. The basis of the assay is that replication in *Tetrahymena* converts plasmid DNA from a Dpn*I*-sensitive to a Dpn*I*-resistant form. With the help of this assay, I isolated a 6.8 kb DNA sequence, called TtARS1, from a *Tetrahymena* macronuclear library that supports autonomous DNA replication of plasmids. TtARS1 is a non-rDNA replicon that was identified from a *Tetrahymena* macronuclear library. Further, I have functionally analyzed this replicon by genetic dissection.

Gene prediction of TtARS1 revealed two hypothetical coding sequences and two intergenic regions. A large coding sequence in the center of TtARS1 was predicted on the basis of its similarity to a hypothetical kinase protein in *Plasmodium*. The second putative coding sequence is partially present in the episomal TtARS1 and corresponds to a hypothetical protein, which is similar to a predicted yeast helicase. There is a 1.2 kb intergenic (non-coding) region between the two predicted genes. Approximately half of a smaller intergenic region of 600 bp is included in the TtARS1 clone.

Functional analysis of non-rDNA replicons

Genetic dissection of TtARS1 revolved on the ability to assess whether TtARS1 derivatives can replicate autonomously in *Tetrahymena*. Interestingly, this functional analysis revealed that TtARS1 in effect contains two independent replication origins, a 966 bp TtARS1-A and a 1.2 kb TtARS1-B. TtARS1-B is completely contained within the predicted larger intergenic region. This is similar to all eukaryotic replicons identified thus far, which reside in non-coding or intergenic DNA segments. Between the two replicons, TtARS1-A was selected for analysis because of its smaller size (966 bp vs. 1.2 kb). Extensive genetic dissection of TtARS1-A mapped the functional origin

to a 700 bp minimal segment that replicates as an episome in *Tetrahymena* (L8 Δ700-800). Further deletion into TtARS1-A (L8 Δ650-750) generates transformants that initially replicate the episome, but gradually lose the plasmid over time. This implies that the sequences contained in the L8 Δ650-750 deletion derivative are sufficient to support ectopic replication activity, but are not able to maintain the long term activity. This may be due to the deletion of cis-acting determinants that are required for replication or a diminished ability to transmit daughter chromosomes to progeny. The former prediction is additionally promoted by the fact that L8 Δ600-700 fails to replicate at all in *Tetrahymena*. Secondarily, plasmids with similar or larger insert sizes fail to replicate at all, so the size of the clone cannot be the rate limiting determinant for origin activity.

S. cerevisiae has served as the pioneering model for eukaryotic DNA replication studies. The replicons in *S. cerevisiae* have a modular structure with discrete cis-acting determinants that are necessary for replicon function (8). Moreover, there are consensus sequences that are conserved between all the replicons in *S. cerevisiae*. In sharp contrast to this, *S. pombe* replicons have a more stochastic nature, where replication initiates at degenerate AT-rich sequences and there is no consensus sequence for replication initiation (14, 19). The above stated genetic dissection of TtARS1 indicates the presence of cis-acting determinants that are necessary for replicon activity. This is especially supported by the fact that some deletion constructs of TtARS1 are completely efficient in supporting replication and others are only partially competent and still others that don't function at all. A very encouraging observation is that not all derivatives of TtARS1 function as *Tetrahymena* replicons. The fact that some deletion constructs fail

to support replication signifies that *Tetrahymena* can be used to identify and functionally dissect replicons at a level of detail that is not currently practical in mammals and *Drosophila*.

To further identify the cis-acting determinants of TtARS1-A, new deletion derivatives (Nhe Δ200-400, Nhe Δ400-600 and Nhe Δ600-4800) have been constructed that contain internal deletions of 200 bp (see figure 2.2). Experiments are currently in progress to study the maintenance of these constructs as episomes in *Tetrahymena*. This will provide further information on the structure of TtARS1 replicon. Interestingly, the functional derivative of TtARS1-A, L8 Δ700-800, contains 300 bp of intergenic region and 400 bp of predicted genic region. This observation is unique since it is the first example in which coding sequences contribute to origin function. For absolute proof, we must determine whether the putative coding sequence is transcribed or not. Ongoing experiments using northern blot analysis should resolve this issue.

Compound replicons in eukaryotes

I mentioned above that TtARS1 contains two distinct replicons within itself, designated TtARS1-A and TtARS1-B. The presence of two discrete replicons within a small region of DNA (6.8 kb) is somewhat atypical and it leads to some speculations. Eukaryotes have very large genome size compared to prokaryotes, and as the rule the cell is compelled to replicate its entire DNA within a small window comprising S phase. Therefore, to deal with this, eukaryotic chromosomes contain multiple origins. The physical proximity of two replication origins is observed in the human β-globin locus. The initiation region in the β-globin replicon contains two adjacent, redundant

replicators, each of which has been shown to initiate replication at ectopic sites (156). Each of the two replicators consists of discrete, non-redundant sequences, which are necessary for replication initiation. Another example of closely spaced initiation regions is the yeast HMR-E region, which consists of at least three independent sub-regions that had the capacity to initiate replication (157). Such compound replicons are often seen in the yeast genome, where the two origins, *ARS101* and *ARS310*, contain multiple, redundant binding sites for ORC (158). Alteration of a single binding site is not sufficient to inactivate the replicons.

Origin usage

The presence of two non-rDNA replicons, TtARS1-A and TtARS1-B, close to each other allows us to speculate on the usage of preferred origins in eukaryotes. In the *GNAI3* locus of Chinese hamster cells, there are multiple replication origins. A frequently used initiation site is about 300 bp downstream of the gene; however there are several other scattered cryptic origins (136, 151). High concentrations of nucleotides results in initiation from the primary origin, whereas reduced nucleotide pools increase initiation events from the cryptic origins. Thus, nucleotide supply modulates the efficiency of active origins and controls activation of dormant origins. Therefore, it is plausible that TtARS1-A and TtARS1-B are both functional replicons, but their activity is differentially regulated, such that they fire under different physiological conditions.

Another probability can be that in the endogenous chromosome only one of the two replicons functions as an origin and the other one is silenced. This might be due to the presence of enhancer sequences in the chromosome that have been excluded in the

episomal TtARS1 fragment. Such long distance effects are often observed in complex replicons of higher eukaryotes like *Drosophila* chorion gene locus where replication enhancer sequences are distributed over a region of 55 kb (159, 160). Another thought is that the chromatin structure of TtARS1 in the chromosome renders an open configuration to one origin and prevents replication from another. Now that the DNA environment is different, both fire in equal capacities. This type of chromatin effects are speculated to occur in the human beta globin replication origin where locus control region that is located about 50 kb away from the origin plays a role in rendering the chromatin in an open state for initiation (29).

Replicator dominance

Prior to this study, the only replicon that had been studied in *Tetrahymena* was the one present in the rDNA minichromosome. The rDNA undergoes locus specific amplification to 10000C within a single S phase during macronuclear development, as compared to non-rDNA chromosomes that are endoreplicated to 45C. This differential replication of rDNA and non-rDNA chromosomes during development is one of the most interesting molecular processes in *Tetrahymena*. Although the steps in macronuclear development have been elucidated, the main factors involved in rDNA amplification are not known. There is a speculation that the rDNA might contain several cis-acting determinants for amplification. Prior deletion studies may have removed too few of these determinants to observe an amplification defect. Alternatively, the non-rDNA chromosomes may have sequence specific information that represses their

amplification. One of the ways to resolve this is to place the rDNA origin of replication adjacent to the non-rDNA origin in an episomal context.

Previous studies have shown that when two copies of the rDNA origin (1.9 kb 5' NTS) are in an episome, the plasmid can maintain itself as a circular episome in *Tetrahymena* (58). For this purpose, two tandem copies of the rDNA 5' NTS were cloned into a plasmid containing the non-rDNA TtARS1-A replicon. Figure 4.1 shows a graphical representation of these chimeric constructs. As a control, the rDNA origin (2 x 5'NTS) was also cloned adjacent to a fragment from the TtARS1 region that fails to support replication in *Tetrahymena* (C6). These plasmids have been transformed into *Tetrahymena* and paromomycin-resistant cultures are being propagated for molecular analysis. TtARS1 and another plasmid containing two copies of the rDNA 5'NTS in pUC vector backbone were included as experimental controls.

There are a few speculations on the result that may be obtained from this experiment. One likelihood is that the juxtaposition of rDNA with the non-rDNA replicon results in firing of the non-rDNA replicon to a high copy number. If this occurs, then this strategy can be used to identify the minimal segment of rDNA specific for amplification. Another thought is that the non-rDNA replicons contain sequences that prevent its amplification. In an episomal context, this region may be sufficient in reducing the copy number of the plasmid carrying the rDNA and non-rDNA origins to a level defined by the non-rDNA replicon. With this approach, the sequence determinant for gene amplification can be analyzed. A gain-of-function study can be done by cloning

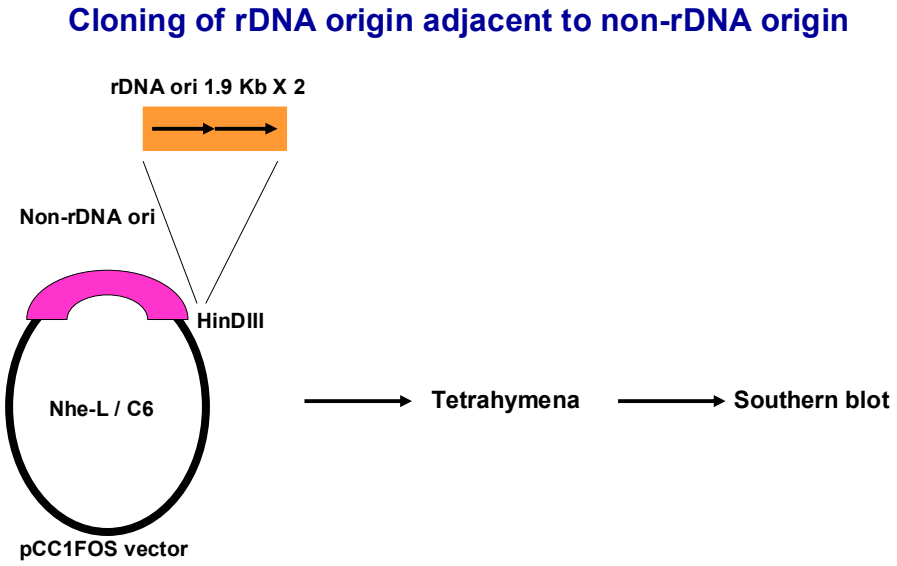


Figure 4.1 Strategy for the study of replicator dominance between r-DNA and non-rDNA replicons.

Two tandem copies of the rDNA origin (5' NTS) was cloned into HinDIII site of two deletion derivatives of TtARS1. One construct was Nhe-L which is functional as an ARS in *Tetrahymena* and the other deletion derivative is C6, which fails to support replication in *Tetrahymena*. The newly generated constructs are designated Chimera 1 and Chimera 2 respectively. *Tetrahymena* co-transformants of the chimeric molecules are generated for Southern blot analysis.

the rDNA segment (amplification determinant) adjacent to the TtARS1-A replicon and testing for the ability to support gene amplification.

Cis-acting determinants for replication initiation: Is there sequence conservation

In this study, two non-rDNA replicons have been characterized with respect to the cis-acting determinants. However, the sites for interaction with replication proteins are not known. In an effort to identify specific sequences or sites that might be binding sites for trans-acting factors in *Tetrahymena*, a comparative analysis between the rDNA and non-rDNA replicon was done using LALIGN software. TtARS1 does not contain any match to the rDNA cis-acting elements such as the type I elements or the pause site elements. There were about sixteen sequences of seven to nine bp that were identical between the rDNA and non-rDNA origins. An attempt was made to look for identical sequences between the two non-rDNA replicons, which revealed twelve identical stretches of six to nine bp. Analysis of two critical segments of 200 bp at the left and right ends of L8 derivative (functional as autonomous replicon in *Tetrahymena*) provided three identical stretches of seven to nine bp. However, identical sequences of eight to ten basepairs were detected through LALIGN analysis of any two arbitrary non-coding *Tetrahymena* sequences of similar lengths (1.0 kb). Thus, with this analysis it was difficult to conclude whether the similar stretches between the origins were meaningful or not as sites for protein binding. An attempt at studying the AT skewness of predicted intergenic regions in TtARS1 as well as the *Tetrahymena* genome was done for fine mapping of the origin. This analysis revealed no significant difference in AT skew between genic and intergenic regions. To this date, only three replicons have been

identified and characterized in *Tetrahymena,* which include the rDNA replicon and the two non-rDNA replicons (TtARS1-A and TtARS1-B). To identify consensus sequences that act as binding sites for replication factors, more origins must be identified and the ARS isolation strategy is a good starting point for this purpose.

TtARS1 replicon was identified based on its ability to support autonomous replication of plasmids. A key question that remains to be answered is whether TtARS1 functions as an origin in the chromosomal context. Of all the origins that have been identified to support ectopic replication in *S. cerevisiae*, only few function as replicons in the native chromosome. This is primarily due to the epigenetic factors such as chromatin environment, DNA methylation and acetylation, amongst others that control initiation of replication (125), or due to the existence of more proficient origins in close proximity. To determine if TtARS1 functions as a chromosomal replicon, two-dimensional gel electrophoresis of replicating intermediates can be used. During the course of this study, an attempt at 2D-gel of TtARS1 was made, but it yielded no result due to limited sensitivity of the assay with non-rDNA chromosomes. An alternative method is nascent strand abundance assay, which has been lately used to map replication origins in higher eukaryotes like Chinese hamster and humans (24).

Trans-acting factors for replication initiation

Currently, others in the lab are generating strains that express tagged *Tetrahymena* ORC and MCM subunits. The expression of the tagged proteins, together with chromatin immunoprecipitation (ChIP) experiment will help in elucidating the replication function of TtARS1 in the endogenous chromosome. The fact that rDNA

undergoes selective gene amplification to 10000C and non-rDNA chromosomes are maintained at 45C implies that different proteins may associate with these two classes of replicons. Thus, the ChIP experiments with TtARS1 and rDNA replicon may ultimately lead to the identification of amplification specific proteins/factors. Also, it is plausible that specific proteins are associated with the non-rDNA replicons that suppress their amplification to high copy number. These proteins can be pulled down by chromatin immunoprecipitation with ORC antibody. Two dimensional gel electrophoresis of proteins, followed by mass spectrometry experiments can be performed to determine the nature and the identity of these proteins.

Due to the limited knowledge of replicons in *Tetrahymena*, the study of regulation of replication is difficult. For example, one interesting question is whether ORC subunits are associated with the origins throughout the cell cycle like in *S. cerevisiae*, or does the complex partially dissociate after origins are replicated as observed in the case of mammals (113, 114, 115, 116, 161). Comparison of the temporal association during the cell cycle between TtARS1 and the rDNA origin will help elucidate the mechanisms for the regulation by ORC.

Studies in our lab have identified four single strand binding proteins, TIF1-4 (Type I binding factors), that regulate rDNA replication. TIF1 has been shown to bind both type I elements and pause site elements in the rDNA 5' NTS (65, 68). It has been observed that *Tetrahymena* mutants that are deficient in TIF1 exhibit early firing of the rDNA replication origin, an elongated macronuclear S-phase and delayed cytokinesis (70). This suggests that TIF1 is not just dedicated to rDNA replication, but plays a global

role in the proper execution and/or monitoring of key chromosomal events (70). TIF2 and TIF3 also bind to type I elements, but TIF3 has been proposed to be a negative regulator of rDNA replication (71). TIF4 is an ORC-like complex that binds in an ATP dependent manner to the rDNA origin (69). Whether these factors are exclusively essential for rDNA replication or play a global role in initiation of replication in *Tetrahymena* is unknown. Currently, experiments are in progress to generate tagged TIF1 protein. ChIP experiments using the tagged cell lines will help in understanding if this protein associate with TtARS1 (non-rDNA) replication initiation. Thus the mechanism of the regulation of the rDNA and the non-rDNA replicons can be unearthed.

Application of the technology: Complementation studies with TtARS1

TtARS1 is maintained as an episome in *Tetrahymena*. This property can be exploited for complementation studies using strains that are partially or completely deficient for specific proteins. A loss-of-function that generated a selectable or easily screened phenotype can be complemented by cloning a wild-type gene in TtARS1 plasmid and the restoration of missing function can be studied. This approach has been used in humans, where a human artificial chromosome was used for the transfer of a human marker gene, hypoxanthine guanine phosphoribosyltransferase (HPRT) (162). Mutations in this gene result in Lesch-Nyhan syndrome. With this method, stable artificial minichromosomes were obtained that complemented the $HPRT^-$ phenotype of the host cells. Since, TtARS1-A is small (966 bp) and is easily maintained as an episome in *Tetrahymena*, DNA hybrids in TtARS1-A plasmids could be introduced into mutants

and screened for rescue of mutant phenotype. The shuttle vector assay could then be used to transfer the rescuing plasmid into *E. coli*.

In my thesis research, I identified two non-rDNA replicons (TtARS1-A and TtARS1-B) in the genetically tractable eukaryote, *Tetrahymena thermophila*. TtARS1-A has been localized to a 700 bp functional segment of DNA and it appears to contain discrete cis-acting regulatory determinants. The fact that there are very few well characterized replicons in higher eukaryotes makes this work important. TtARS1 serves as an important member of the family of eukaryotic replicons and most importantly can be used to unearth the mechanism of selective gene amplification in *Tetrahymena*. The development of ARS isolation strategy is of particular significance since it is a reliable method for isolation and genetic dissection of new origins. The identification of TtARS1 will thus help in elucidating in molecular detail the mechanisms that control DNA replication in this model eukaryote.

REFERENCES

1. DePamphilis, M.L. (2005) Cell cycle dependent regulation of the origin recognition complex. *Cell Cycle.,* **4,** 70-9.

2. Bell, S.P. and Dutta, A. (2002) DNA replication in eukaryotic cells. *Annu Rev Biochem.,* **71**, 333-74.

3. Jacob, F., Brenner, S. and Cuzin, F. (1963) On the regulation of DNA synthesis in bacteria: the hypothesis of the replicon. *Cold Spring Harb Symp Quant Biol.,* **263**, 298-300.

4. Bell, S.P. and Stillman, B. (1992) ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature,* **357**, 128-34.

5. Gavin, K.A., Hidaka, M. and Stillman, B. (1995) Conserved initiator proteins in eukaryotes. *Science,* **270**, 1667-71.

6. Huberman, J.A. and Riggs, A.D. (1968) On the mechanism of DNA replication in mammalian chromosomes. *J Mol Biol.,* **32,** 327-41.

7. Williamson, D.H. (1985) The yeast ARS element, six years on: a progress report. *Yeast,* **1,** 1-14.

8. Marahrens, Y. and Stillman, B. (1992) A yeast chromosomal origin of DNA replication defined by multiple functional elements. *Science,* **255,** 817-23.

9. Rao, H., Marahrens, Y. and Stillman, B. (1994) Functional conservation of multiple elements in yeast chromosomal replicators. *Mol Cell Biol.,* **14,** 7643-51.

10. Newlon, C.S. and Theis, J.F. (1993) The structure and function of yeast ARS elements. *Curr Opin Genet Dev.,* **3,** 752-8.

11. Wilmes, G.M. and Bell, S.P. (2002) The B2 element of the *Saccharomyces cerevisiae* ARS1 origin of replication requires specific sequences to facilitate pre-RC formation. *Proc Natl Acad Sci USA,* **99**, 101-6.

12. Walker, S.S., Malik, A.K. and Eisenberg, S. (1990) Analysis of the interactions of functional domains of a nuclear origin of replication from *Saccharomyces cerevisiae*. *Nucleic Acids Res.,* **19,** 6255-62.

13. Sakaguchi, J. and Yamamoto, M. (1982) Cloned ural locus of *Schizosaccharomyces pombe* propagates autonomously in this yeast assuming a polymeric form. *Proc Natl Acad Sci USA.,* **79**, 7819-23.

14. Kim, S.M. and Huberman, J.A. (1998) Multiple orientation-dependent, synergistically interacting, similar domains in the ribosomal DNA replication origin of the fission yeast, *Schizosaccharomyces pombe*. *Mol Cell Biol.,* **18**, 7294-303.

15. Okuno, Y., Satoh, H., Sekiguchi, M. and Masukata, H. (1999) Clustered adenine/thymine stretches are essential for function of a fission yeast replication origin. *Mol Cell Biol.,* **19,** 6699-709.

16. Gomez, M. and Antequera, F. (1999) Organization of DNA replication origins in the fission yeast genome. *Embo J.,* **18**, 5683-90.

17. Chuang, R.Y. and Kelly, T.J. (1999) The fission yeast homologue of Orc4p binds to replication origin DNA via multiple AT-hooks. *Proc Natl Acad Sci USA,* **96**, 2656-61.

18. Lee, J.K., Moon, K.Y., Jiang, Y. and Hurwitz, J. (2001) The *Schizosaccharomyces pombe* origin recognition complex interacts with multiple AT-rich regions of the replication origin DNA by means of the AT-hook domains of the spOrc4 protein. *Proc Natl Acad Sci USA,* **98**, 13589-94.

19. Dai, J., Chuang, R.Y. and Kelly, T.J. (2005) DNA replication origins in the *Schizosaccharomyces pombe* genome. *Proc Natl Acad Sci USA,* **102**, 337-42.

20. Heintz, N.H. and Hamlin, J.L. (1982) An amplified chromosomal sequence that includes the gene for dihydrofolate reductase initiates replication within specific restriction fragments. *Proc Natl Acad Sci USA,* **79**, 4083-7.

21. Leu, T.H. and Hamlin, J.L. (1989) High-resolution mapping of replication fork movement through the amplified dihydrofolate reductase domain in CHO cells by in-gel renaturation analysis. *Mol Cell Biol.,* **9**, 523-31.

22. Anachkova, B. and Hamlin, J.L. (1989) Replication in the amplified dihydrofolate reductase domain in CHO cells may initiate at two distinct sites, one of which is a repetitive sequence element. *Mol Cell Biol.,* **9**, 532-40.

23. Ma, C., Leu, T.H. and Hamlin, J.L. (1990) Multiple origins of replication in the dihydrofolate reductase amplicons of a methotrexate-resistant Chinese hamster cell line. *Mol Cell Biol.,* **10**, 1338-46.

24. Altman, A.L. and Fanning, E. (2001) The Chinese hamster dihydrofolate reductase replication origin beta is active at multiple ectopic chromosomal locations and requires specific DNA sequence elements for activity. *Mol Cell Biol*., **21**, 1098-110.

25. Kobayashi, T., Rein, T. and DePamphilis, M.L. (1998) Identification of primary initiation sites for DNA replication in the hamster dihydrofolate reductase gene initiation zone. *Mol Cell Biol.,* **18**, 3266-77.

26. Caddle, M.S., Lussier, R.H. and Heintz, N.H. (1990) Intramolecular DNA triplexes, bent DNA and DNA unwinding elements in the initiation region of an amplified dihydrofolate reductase replicon. *J Mol Biol.,* **211**, 19-33.

27. Kitsberg, D., Selig, S., Keshet, I. and Cedar, H. (1993) Replication structure of the human beta-globin gene domain. *Nature,* **366**, 588-90.

28. Aladjem, M.I., Groudine, M., Brody, L.L., Dieken, E.S., Fournier, R.E., Wahl, G.M. and Epner, E.M. (1995) Participation of the human beta-globin locus control region in initiation of DNA replication. *Science,* **270**, 815-9.

29. Aladjem, M.I., Rodewald, L.W., Kolman, J.L. and Wahl, G.M. (1998) Genetic dissection of a mammalian replicator in the human beta-globin locus. *Science,* **281**, 1005-9.

30. Simon, I., Tenzen, T., Mostoslavsky, R., Fibach, E., Lande, L., Milot, E., Gribnau, J., Grosveld, F., Fraser, P. and Cedar, H. (2001) Developmental regulation of DNA replication timing at the human beta globin locus. *Embo J.,* **20**, 6150-7.

31. Giacca, M., Zentilin, L., Norio, P., Diviacco, S., Dimitrova, D., Contreas, G., Biamonti, G., Perini, G., Weighardt, F., Riva, S., et al. (1994) Fine mapping of a replication origin of human DNA. *Proc Natl Acad Sci USA.,* **91,** 7119-23.

32. Biamonti, G., Perini, G., Weighardt, F., Riva, S., Giacca, M., Norio, P., Zentilin, L., Diviacco, S., Dimitrova, D. and Falaschi. A. (1992) A human DNA replication origin: localization and transcriptional characterization. *Chromosoma,* **102**, S24-31.

33. Paixao, S., Colaluca, I.N., Cubells, M., Peverali, F.A., Destro, A., Giadrossi, S., Giacca, M., Falaschi, A., Riva, S. and Biamonti, G. (2004) Modular structure of the human lamin B2 replicator. *Mol Cell Biol.,* **24**, 2958-67.

34. Abdurashidova, G., Deganuto, M., Klima, R., Riva, S., Biamonti, G., Giacca, M. and Falaschi, A. (2000) Start sites of bidirectional DNA synthesis at the human lamin B2 origin. *Science,* **287**, 2023-6.

35. Leffak, M. and James, C.D. (1989) Opposite replication polarity of the germ line c-myc gene in HeLa cells compared with that of two Burkitt lymphoma cell lines. *Mol Cell Biol.,* **9**, 586-93.

36. Trivedi, A., Waltz, S.E., Kamath, S. and Leffak, M. (1998) Multiple initiations in the c-myc replication origin independent of chromosomal location. *DNA Cell Biol.,* **17**, 885-96.

37. Dobbs, D.L., Shaiu, W.L. and Benbow, R.M. (1994) Modular sequence elements associated with origin regions in eukaryotic chromosomal DNA. *Nucleic Acids Res.,* **22**, 2479-89.

38. Berberich, S., Trivedi, A., Daniel, D.C., Johnson, E.M. and Leffak, M. (1995) *In vitro* replication of plasmids containing human c-myc DNA. *J Mol Biol.,* **245**, 92-109.

39. Malott, M. and Leffak, M. (1999) Activity of the c-myc replicator at an ectopic chromosomal location. *Mol Cell Biol.,* **19**, 5685-95.

40. Heinzel, S.S., Krysan, P.J., Tran, C.T. and Calos, M.P. (1991) Autonomous DNA replication in human cells is affected by the size and the source of the DNA. *Mol Cell Biol.,* **11**, 2263-72.

41. Spradling, A.C. (1981) The organization and amplification of two chromosomal domains containing *Drosophila* chorion genes. *Cell,* **27**, 193-201.

42. Levine, J. and Spradling, A. (1985) DNA sequence of a 3.8 kilobase pair region controlling *Drosophila* chorion gene amplification. *Chromosoma,* **92**, 136-42.

43. Delidakis, C. and Kafatos, F.C. (1987) Amplification of a chorion gene cluster in *Drosophila* is subject to multiple cis-regulatory elements and to long-range position effects. *J Mol Biol.,* **197**, 11-26.

44. Zhang, H. and Tower, J. (2004) Sequence requirements for function of the *Drosophila* chorion gene locus ACE3 replicator and ori-beta origin elements. *Development*., **131**, 2089-99.

45. Orr-Weaver, T.L., Johnston, C.G. and Spradling, A.C. (1989) The role of ACE3 in *Drosophila* chorion gene amplification. *Embo J.,* **8**, 4153-62.

46. Heck, M.M. and Spradling, A.C. (1990) Multiple replication origins are used during *Drosophila* chorion gene amplification. *J Cell Biol.,* **110**, 903-14.

47. Carminati, J.L., Johnston, C.G. and Orr-Weaver, T.L. (1992) The *Drosophila* ACE3 chorion element autonomously induces amplification. *Mol Cell Biol.,* **12**, 2444-53.

48. de Cicco, D.V. and Spradling, A.C. (1984) Localization of a cis-acting element responsible for the developmentally regulated amplification of *Drosophila* chorion genes. *Cell,* **38**, 45-54.

49. Lu, L., Zhang, H. and Tower, J. (2001) Functionally distinct, sequence-specific replicator and origin elements are required for *Drosophila* chorion gene amplification. *Genes Dev*., **15,** 134-46.

50. Remus, D., Beall, E. L. and Botchan, M.R. (2004) DNA topology, not DNA sequence, is a critical determinant for *Drosophila* ORC-DNA binding. *Embo J.,* **23**, 897-907.

51. Smith, J.G. and M.P. Calos (1995) Autonomous replication in *Drosophila melanogaster* tissue culture cells. *Chromosoma,* **103**, 597-605.

52. Karrer, K.M. and J.G. Gall (1976) The macronuclear ribosomal DNA of *Tetrahymena pyriformis* is a palindrome. *J Mol Biol.,* **104**, 421-53.

53. Engberg, J., Andersson, P., Leick, V. and Collins, J. (1976) Free ribosomal DNA molecules from *Tetrahymena pyriformis* GL are giant palindromes. *J Mol Biol.,* **104**, 455-70.

54. Yao, M.C. (1981) Ribosomal RNA gene amplification in *Tetrahymena* may be associated with chromosome breakage and DNA elimination. *Cell*, **24**, 765-74.

55. Yasuda, L.F. and Yao, M.C. (1991) Short inverted repeats at a free end signal large palindromic DNA formation in *Tetrahymena*. *Cell*, **67**, 505-16.

56. Butler, D.K., Yasuda, L.E. and Yao, M.C. (1995) An intramolecular recombination mechanism for the formation of the rRNA gene palindrome of *Tetrahymena thermophila*. *Mol Cell Biol.,* **15**, 7117-26.

57. Amin, A.A. and R.E. Pearlman (1986) *In vitro* deletion analysis of ARS elements spanning the replication origin in the 5' non-transcribed spacer of *Tetrahymena thermophila* ribosomal DNA. *Nucleic Acids Res.,* **14**, 2749-62.

58. Pan, W.J. and Blackburn, E.H. (1995) Tandem repeats of the 5' non-transcribed spacer of *Tetrahymena* rDNA function as high copy number autonomous replicons in the macronucleus but do not prevent rRNA gene dosage regulation. *Nucleic Acids Res*., **23**, 1561-9.

59. Larson, D.D., Blackburn, E.H., Yaeger, P.C. and Orias, E. (1986) Control of rDNA replication in *Tetrahymena* involves a cis-acting upstream repeat of a promoter element. *Cell,* **47**, 229-40.

60. Reischmann, K.P., Zhang, Z. and Kapler, G.M. (1999) Long range cooperative interactions regulate the initiation of replication in the *Tetrahymena thermophila* rDNA minichromosome. *Nucleic Acids Res.,* **27**, 3079-89.

61. Challoner, P.B., Amin, A.A., Pearlman, R.E. and Blackburn, E.H. (1985) Conserved arrangements of repeated DNA sequences in nontranscribed spacers of ciliate ribosomal RNA genes: evidence for molecular coevolution. *Nucleic Acids Res.,* **13**, 2661-80.

62. Gallagher, R.C. and Blackburn, E.H. (1998) A promoter region mutation affecting replication of the *Tetrahymena* ribosomal DNA minichromosome. *Mol Cell Biol.,* **18**, 3021-33.

63. Kapler, G.M., Orias, E. and Blackburn, E.H. (1994) *Tetrahymena thermophila* mutants defective in the developmentally programmed maturation and maintenance of the rDNA minichromosome. *Genetics*, **137**, 455-66.

64. MacAlpine, D.M., Zhang, Z. and Kapler, G.M. (1997) Type I elements mediate replication fork pausing at conserved upstream sites in the *Tetrahymena thermophila* ribosomal DNA minichromosome. *Mol Cell Biol*., **17**, 4517-25.

65. Saha, S., Nicholson, A. and Kapler, G.M. (2001) Cloning and biochemical analysis of the *Tetrahymena* origin binding protein TIF1: competitive DNA binding *in vitro* and *in vivo* to critical rDNA replication determinants. *J Biol Chem.,* **276**, 45417-26.

66. Zhang, Z., Macalpine, D.M. and Kapler, G.M. (1997) Developmental regulation of DNA replication: replication fork barriers and programmed gene amplification in *Tetrahymena thermophila*. *Mol Cell Biol.,* **17**, 6147-56.

67. Umthun, A.R., Hou, Z., Sibenaller, Z.A., Shaiu, W.L. and Dobbs, D.L. (1994) Identification of DNA-binding proteins that recognize a conserved type I repeat sequence in the replication origin region of *Tetrahymena* rDNA. *Nucleic Acids Res.,* **22**, 4432-40.

68. Saha, S. and G. M. Kapler (2000) Allele-specific protein-DNA interactions between the single-stranded DNA-binding protein, ssA-TIBF, and DNA replication determinants in *Tetrahymena*. *J Mol Biol.,* **295**, 423-39.

69. Mohammad, M., York, R. D., Hommel, J. and Kapler, G.M. (2003) Characterization of a novel origin recognition complex-like complex: implications for DNA recognition, cell cycle control, and locus-specific gene amplification. *Mol Cell Biol.,* **23**, 5005-17.

70. Morrison, T.L., Yakisich, J.S., Cassidy-Hanley, D. and Kapler, G.M. (2005) TIF1 Represses rDNA replication initiation, but promotes normal S phase progression and chromosome transmission in *Tetrahymena*. *Mol Biol Cell.,* **16**, 2624-35.

71. Mohammad, M., Saha, S. and Kapler, G.M. (2000) Three different proteins recognize a multifunctional determinant that controls replication initiation, fork arrest and transcription in *Tetrahymena*. *Nucleic Acids Res.,* **28**, 843-51.

72. Mohammad, M. and Kapler, G.M. (2004) personal communication. Dept. of Medical Biochemistry and Genetics, Texas A&M University, College Station, Texas.

73. Aparicio, O.M., Weinstein, D.M. and Bell, S.P. (1997) Components and dynamics of DNA replication complexes in *S. cerevisiae*: redistribution of MCM proteins and Cdc45p during S phase. *Cell,* **91**, 59-69.

74. Mendez, J. and Stillman, B. (2000) Chromatin association of human origin recognition complex, cdc6, and minichromosome maintenance proteins during the cell cycle: assembly of prereplication complexes in late mitosis. *Mol Cell Biol*., **20**, 8602-12.

75. Tanaka, T., Knapp, D. and Nasmyth, K. (1997) Loading of an Mcm protein onto DNA replication origins is regulated by Cdc6p and CDKs. *Cell*, **90**, 649-60.

76. Rao, H. and Stillman, B. (1995) The origin recognition complex interacts with a bipartite DNA binding site within yeast replicators. *Proc Natl Acad Sci USA,* **92**, 2224-8.

77. Lee, D.G. and Bell, S.P. (1997) Architecture of the yeast origin recognition complex bound to origins of DNA replication. *Mol Cell Biol.,* **17**, 7159-68.

78. Diffley, J.F. and Cocker, J.H. (1992) Protein-DNA interactions at a yeast replication origin. *Nature,* **357**, 169-72.

79. Nurse, P., Thuriaux, P. and Nasmyth, K. (1976) Genetic control of the cell division cycle in the fission yeast *Schizosaccharomyces pombe*. *Mol Gen Genet.,* **146**, 167-78.

80. Weinreich, M. and Stillman, B. (1999) Cdc7p-Dbf4p kinase binds to chromatin during S phase and is regulated by both the APC and the RAD53 checkpoint pathway. *Embo J.,* **18**, 5334-46.

81. Tanaka, S. and Diffley, J.F. (2002) Interdependent nuclear accumulation of budding yeast Cdt1 and Mcm2-7 during G1 phase. *Nat Cell Biol.,* **4**, 198-207.

82. Li, A. and Blow, J.J. (2005) Cdt1 downregulation by proteolysis and geminin inhibition prevents DNA re-replication in *Xenopus*. *Embo J*., **24**, 395-404.

83. Maine, G.T., Sinha, P. and Tye, B.K. (1984) Mutants of *S. cerevisiae* defective in the maintenance of minichromosomes. *Genetics,* **106**, 365-85.

84. Chong, J.P., Mahbubani, H.M., Khoo, C.Y. and Blow, J.J. (1995) Purification of an MCM-containing complex as a component of the DNA replication licensing system. *Nature,* **375**, 418-21.

85. Lee, J.K. and Hurwitz, J. (2000) Isolation and characterization of various complexes of the minichromosome maintenance proteins of Schizosaccharomyces pombe. *J Biol Chem*, **275**, 18871-8.

86. You, Z., Komamura, Y. and Ishimi, Y. (1999) Biochemical analysis of the intrinsic Mcm4-Mcm6-mcm7 DNA helicase activity. *Mol Cell Biol.,* **19**, 8003-15.

87. Ishimi, Y. (1997) A DNA helicase activity is associated with an MCM4, -6, and -7 protein complex. *J Biol Chem.,* **272**, 24508-13.

88. Merchant, A.M., Kawasaki, Y., Chen, Y., Lei, M. and Tye ,B.K. (1997) A lesion in the DNA replication initiation factor Mcm10 induces pausing of elongation forks through chromosomal replication origins in *Saccharomyces cerevisiae*. *Mol Cell Biol.,* **17**, 3261-71.

89. Ricke, R.M. and A.K. Bielinsky (2004) Mcm10 regulates the stability and chromatin association of DNA polymerase-alpha. *Mol Cell,* **16**, 173-85.

90. Wohlschlegel, J.A., Dhar, S.K., Prokhorova, T.A., Dutta, A. and Walter, J.C. (2002) *Xenopus* Mcm10 binds to origins of DNA replication after Mcm2-7 and stimulates origin binding of Cdc45. *Mol Cell,* **9**, 233-40.

91. Gozuacik, D., Chami, M., Lagorce, D., Faivre, J., Murakami, Y., Poch, O., Biermann, E., Knippers, R., Brechot, C. and Paterlini-Brechot, P. (2003) Identification and functional characterization of a new member of the human Mcm protein family: hMcm8. *Nucleic Acids Res.,* **31**, 570-9.

92. Maiorano, D., Cuvier, O., Danis, E. and Mechali, M. (2005). MCM8 is an MCM2-7-related protein that functions as a DNA helicase during replication elongation and not initiation. *Cell,* **120**, 315-28.

93. Bousset, K. and Diffley, J.F. (1998) The Cdc7 protein kinase is required for origin firing during S phase. *Genes Dev.,* **12**, 480-90.

94. Donaldson, A.D., Fangman, W.L. and Brewer, B.J. (1998) Cdc7 is required throughout the yeast S phase to activate replication origins. *Genes Dev.,* **12**, 491-501.

95. Masai, H., Matsui, E., You, Z., Ishimi, Y., Tamai, K. and Arai, K. (2000) Human Cdc7-related kinase complex. *In vitro* phosphorylation of MCM by concerted actions of Cdks and Cdc7 and that of a criticial threonine residue of Cdc7 by Cdks. *J Biol Chem.,* **275**, 29042-52.

96. Zou, L. and Stillman, B. (2000) Assembly of a complex containing Cdc45p, replication protein A, and Mcm2p at replication origins controlled by S-phase cyclin-dependent kinases and Cdc7p-Dbf4p kinase. *Mol Cell Biol.,* **20**, 3086-96.

97. Jares, P. and Blow, J.J. (2000) *Xenopus* cdc7 function is dependent on licensing but not on XORC, XCdc6, or CDK activity and is required for XCdc45 loading. *Genes Dev.,* **14**, 1528-40.

98. Zou, L., Mitchell, J. and Stillman, B. (1997) CDC45, a novel yeast gene that functions with the origin recognition complex and Mcm proteins in initiation of DNA replication. *Mol Cell Biol.,* **17**, 553-63.

99. Aparicio, O.M., Stout, A.M. and Bell, S.P. (1999) Differential assembly of Cdc45p and DNA polymerases at early and late origins of DNA replication. *Proc Natl Acad Sci USA,* **96**, 9130-5.

100. Tercero, J.A., Labib, K. and Diffley, J.F. (2000) DNA synthesis at individual replication forks requires the essential initiation factor Cdc45p. *Embo J.,* **19**, 2082-93.

101. Tanaka, T. and Nasmyth, K. (1998) Association of RPA with chromosomal replication origins requires an Mcm protein, and is regulated by Rad53, and cyclin- and Dbf4-dependent kinases. *Embo J.,* **17**, 5182-91.

102. Mimura, S., Masuda, T., Matsui, T. and Takisawa, H. (2000) Central role for cdc45 in establishing an initiation complex of DNA replication in *Xenopus* egg extracts. *Genes Cells,* **5**, 439-52.

103. Nguyen, V.Q., Co, C. and Li, J.J. (2001) Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. *Nature,* **411**, 1068-73.

104. Elsasser, S., Lou, F., Wang, B., Campbell, J.L. and Jong, A. (1996) Interaction between yeast Cdc6 protein and B-type cyclin/Cdc28 kinases. *Mol Biol Cell,* 7, 1723-35.

105. Jallepalli, P.V., Brown, G.W., Muzi-Falconi, M., Tien, D. and Kelly, T.J. (1997) Regulation of the replication initiator protein p65cdc18 by CDK phosphorylation. *Genes Dev*., **11**, 2767-79.

106. Petersen, B.O., Lukas, J., Sorensen, C.S., Bartek, J. and Helin, K. (1999) Phosphorylation of mammalian CDC6 by cyclin A/CDK2 regulates its subcellular localization. *Embo J.,* **18**, 396-410.

107. Jiang, W., Wells, N.J. and Hunter, T. (1999) Multistep regulation of DNA replication by Cdk phosphorylation of HsCdc6. *Proc Natl Acad Sci USA,* **96,** 6193-8.

108. Yan, H., Merchant, A.M. AND Tye, B.K. (1993) Cell cycle-regulated nuclear localization of MCM2 and MCM3, which are required for the initiation of DNA synthesis at chromosomal replication origins in yeast. *Genes Dev.,* **7**, 2149-60.

109. Lei, M., Kawasaki, Y. and Tye, B.K. (1996) Physical interactions among Mcm proteins and effects of Mcm dosage on DNA replication in *Saccharomyces cerevisiae*. *Mol Cell Biol.,* **16**, 5081-90.

110. Hendrickson, M., Madine, M., Dalton, S. and Gautier, J. (1996) Phosphorylation of MCM4 by cdc2 protein kinase inhibits the activity of the minichromosome maintenance complex. *Proc Natl Acad Sci USA,* **93**, 12223-8.

111. Hua, X. H., Yan, H. and Newport, J. (1997) A role for Cdk2 kinase in negatively regulating DNA replication during S phase of the cell cycle. *J Cell Biol.,* **137**, 183-92.

112. Fujita, M., Yamada, C., Tsurumi, T., Hanaoka, F., Matsuzawa, K. AND Inagaki, M. (1998) Cell cycle- and chromatin binding state-dependent of human MCM heterohexameric complexes, a role for cdc2 kinase. *J Biol Chem.,* **273**, 17095-101.

113. Fujita, M., Ishimi, Y., Nakamura, H., Kiyono, T. and Tsurumi, T. (2002) Nuclear organization of DNA replication initiation proteins in mammalian cells. *J Biol Chem.,* **277**, 10354-61.

114. Mendez, J., Zou-Yang, X. H., Kim, S.Y., Hidaka, M., Tansey, W.P. and Stillman, B. (2002) Human origin recognition complex large subunit is degraded by ubiquitin-mediated proteolysis after initiation of DNA replication. *Mol Cell,* **9**, 481-91.

115. Asano, M. and Wharton, R.P. (1999) E2F mediates developmental and cell cycle regulation of ORC1 in *Drosophila*. *Embo J.,* **18**, 2435-48.

116. Royzman, I., Austin, R.J., Bosco, G., Bell, S.P. and Orr-Weaver, T.L. (1999) ORC localization in *Drosophila* follicle cells and the effects of mutations in dE2F and dDP. *Genes Dev.,* **13**, 827-40.

117. Romanowski, P., Madine, M.A., Rowles, A., Blow, J.J. and Laskey, R.A. (1996) The *Xenopus* origin recognition complex is essential for DNA replication and MCM binding to chromatin. *Curr Biol.,* **6**, 1416-25.

118. Nakayasu, H. and Berezney, R. (1989) Mapping replicational sites in the eukaryotic cell nucleus. *J Cell Biol*., **108**, 1-11.

119. Berezney, R. and Wei, X. (1998) The new paradigm: integrating genomic function and nuclear architecture. *J Cell Biochem Suppl.,* **30,** 238-42.

120. Stambrook, P.J. and Flickinger, R.A. (1970) Changes in chromosomal DNA replication patterns in developing frog embryos. *J Exp Zool*., **174**, 101-13.

121. Taljanidisz, J., Popowski, J. and Sarkar, N. (1989) Temporal order of gene replication in Chinese hamster ovary cells. *Mol Cell Biol.,* **9**, 2881-9.

122. Goldman, M.A., Holmquist, G.P., Gray, M.C., Caston, L.A. and Nag, A. (1984) Replication timing of genes and middle repetitive sequences. *Science,* **224**, 686-92.

123. Yates, J. L., Camiolo, S.M. and Bashaw, J.M. (2000) The minimal replicator of Epstein-Barr virus oriP. *J Virol.,* **74**, 4512-22.

124. Spradling, A.C., de Cicco, D.V., Wakimoto, B.T., Levine, J.F., Kalfayan, L.J. and Cooley, L. (1987) Amplification of the X-linked Drosophila chorion gene cluster requires a region upstream from the s38 chorion gene.*EMBO J*., **4**,1045-53.

125. Zhou, J., Chau, C.M., Deng, Z., Shiekhattar, R., Spindler, M.P., Schepers, A. and Lieberman, P.M. (2005) Cell cycle regulation of chromatin at an origin of DNA replication. *Embo J.,* **24**, 1406-17.

126. Aggarwal, B.D. and Calvi, B.R. (2004) Chromatin regulates origin activity in *Drosophila* follicle cells. *Nature,* **430**, 372-6.

127. Vogelauer, M., Rubbi, L., Lucas, I., Brewer, B.J. and Grunstein, M. (2002) Histone acetylation regulates the time of replication origin firing. *Mol Cell.,* **10**, 1223-33.

128. Bird, A.P. (1986) CpG-rich islands and the function of DNA methylation. *Nature,* **321**, 209-13.

129. Larsen, F., Gundersen, G., Lopez, R. and Prydz, H. (1992) CpG islands as gene markers in the human genome. *Genomics,* **13**, 1095-107.

130. Rein, T., Zorbas, H. and DePamphilis, M.L. (1997) Active mammalian replication origins are associated with a high-density cluster of mCpG dinucleotides. *Mol Cell Biol.,* **17**, 416-26.

131. Delgado, S., Gomez, M., Bird, A. and Antequera, F. (1998) Initiation of DNA replication at CpG islands in mammalian chromosomes. *Embo J.,* **17**, 2426-35.

132. Harvey, K.J. and Newport, J. (2003) CpG methylation of DNA restricts prereplication complex assembly in *Xenopus* egg extracts. *Mol Cell Biol.,* **23**, 6769-79.

133. Rein, T., Kobayashi, T., Malott, M., Leffak, M. and DePamphilis, M.L. (1999) DNA methylation at mammalian replication origins. *J Biol Chem.,* **274**, 25792-800.

134. Zappulla, D.C., Sternglanz, R. and Leatherwood, J. (2002) Control of replication timing by a transcriptional silencer. *Curr Biol.,* **12**, 869-75.

135. Danis, E., Brodolin, K., Menut, S., Maiorano, D., Girard-Reydet, C. and Mechali, M. (2004) Specification of a DNA replication origin by a transcription complex. *Nat Cell Biol.,* **6**, 721-30.

136. Anglana, M., Apiou, F., Bensimon, A. and Debatisse, M. (2003) Dynamics of DNA replication in mammalian somatic cells: nucleotide pool modulates origin choice and interorigin spacing. *Cell,* **114**, 385-94.

137. Burhans, W.C., Selegue, J.E. and Heintz, N.H. (1986) Isolation of the origin of

replication associated with the amplified Chinese hamster dihydrofolate reductase domain. *Proc Natl Acad Sci USA,* **83**, 7790-4.

138. Russev, G. and Vassilev, L. (1982) Isolation of a DNA fraction from Ehrlich ascites tumour cells containing the putative origin of replication. *J Mol Biol.,* **161**, 77-87.

139. Kaufmann, G., Zannis-Hadjopoulos, M. and Martin, R.G. (1985) Cloning of nascent monkey DNA synthesized early in the cell cycle. *Mol Cell Biol.,* **5**, 721-7.

140. Stinchcomb, D. T., Struhl, K. and Davis, R.W. (1979) Isolation and characterization of a yeast chromosomal replicator. *Nature,* **282**, 39-43.

141. Asai, D.J. and Forney, J.D. (2000) *Methods in Cell Biology*. **Volume 62,** *Tetrahymena thermophila*. 1st Edn. Academic Press, New York.

142. Yao, M. C., Zheng, K. and Yao, C.H. (1987) A conserved nucleotide sequence at the sites of developmentally regulated chromosomal breakage in *Tetrahymena*. *Cell,* **48**, 779-88.

143. Fan, Q. and Yao, M. (1996) New telomere formation coupled with site-specific chromosome breakage in *Tetrahymena thermophila*. *Mol Cell Biol.,* **16**, 1267-74.

144. Nanney, D. L., Meyer, E. B. and Portnoy, S. (1980) Perturbance analysis of nuclear determination in *Tetrahymena*: effects of nutrition, cell extracts, and CaCl2 on A/B hybrids. *Differentiation,* **16**, 49-60.

145. Orias, E., Hashimoto, N., Chau, M.F. and Higashinakagawa, T. (1991) PCR amplification of *Tetrahymena* rDNA segments starting with individual cells. *J Protozool.,* **38**, 306-11.

146. Gall, J.G. (1974) Free ribosomal RNA genes in the macronucleus of *Tetrahymena*. *Proc Natl Acad Sci USA,* **71**, 3078-81.

147. Yao, M.C. and Yao, C.H. (1991) Transformation of *Tetrahymena* to cycloheximide resistance with a ribosomal protein gene through sequence replacement. *Proc Natl Acad Sci USA,* **88**, 9493-7.

148. Yao, M.C. and Gorovsky, M.A. (1974) Comparison of the sequences of macro- and micronuclear DNA of *Tetrahymena pyriformis*. *Chromosoma,* **48**, 1-18.

149. Bruns, P.J. and Cassidy-Hanley, D. (2000) Biolistic transformation of macro- and micronuclei. *Methods Cell Biol.* **62,** 501-12.

150. The Institute for Genomic Research (TIGR). (1992) *Tetrahymena thermophila* Genome Project accessed on December 11, 2003. http://www.tigr.org.

151. Aladjem, M.I. and Fanning, E. (2004) The replicon revisited: an old model learns new tricks in metazoan chromosomes. *EMBO Rep.,* **5**, 686-91.

152. Robinson, N.P. and Bell, S.D. (2005) Origins of DNA replication in the three domains of life. *Febs J.,* **272**, 3757-66.

153. Yaeger, P.C., Orias, E., Shaiu, W.L., Larson, D.D. and Blackburn, E.H. (1989) The replication advantage of a free linear rRNA gene is restored by somatic recombination in *Tetrahymena thermophila*. *Mol Cell Biol*. **9,** 452-60.

154. Kearsey, S. (1983) Analysis of sequences conferring autonomous replication in baker's yeast. *Embo J.,* **2**, 1571-5.

155. Krysan, P.J., Smith, J.G. and Calos, M.P. (1993) Autonomous replication in human cells of multimers of specific human and bacterial DNA sequences. *Mol Cell Biol.,* **13**, 2688-96.

156. Wang, L., Lin, C.M., Brooks, S., Cimbora, D., Groudine, M. and Aladjem, M.I. (2004) The human beta-globin replication initiation region consists of two modular independent replicators. *Mol Cell Biol.,* **24**, 3373-86.

157. Hurst, S.T. and Rivier, D.H. (1999) Identification of a compound origin of replication at the HMR-E locus in *Saccharomyces cerevisiae*. *J Biol Chem.,* **274**, 4155-9.

158. Theis, J.F. and Newlon, C.S. (2001) Two compound replication origins in *Saccharomyces cerevisiae* contain redundant origin recognition complex binding sites. *Mol Cell Biol.,* **21**, 2790-801.

159. Delidakis, C. and Kafatos, F.C. (1989) Amplification enhancers and replication origins in the autosomal chorion gene cluster of *Drosophila*. *Embo J.,* **8,** 891-901.

160. Gilbert, D.M. (2001) Making sense of eukaryotic DNA replication origins. *Science,* **294,** 96-100.

161. DePamphilis, M.L. (2003) The 'ORC cycle': a novel pathway for regulating eukaryotic DNA replication. *Gene,* **310**, 1-15.

162. Mejia, J.E. and Larin, Z. (2000) The assembly of large BACs by *in vivo* recombination. *Genomics,* **70**, 165-70.

VITA

Shibani Datta
Jai Krishna Cooperative Housing Society
Radha-2
Chhedenagar
Chembur
Mumbai 400089
India

Email: shibanidatta79@yahoo.com

EDUCATION

University of Mumbai, India, B.S., Microbiology, May 2000, **GPA 3.9**
University of Mumbai, India, M.S., Microbiology, May 2002, **GPA 3.9**
Texas A&M University, M.S., Genetics, December 2005, **GPA 3.84**

PROFESSIONAL EXPERIENCE

2001-2001: Research Intern at the Bhabha Atomic Research Centre, Mumbai, India.
Conducted research on studying the differences in radiation response in repair deficient strains of yeast, Saccharomyces cerevisiae

2003-2004: Teaching assistant in *Genetics,* Gene 301, Fall 2003, for Dr.
Dr. Ellisson, Biochemistry and Biophysics, Texas A&M University.

2002-2005: Research Assistant at Texas A&M University, in Dr. Geoffrey Kapler's laboratory.