Introduction

One universal feature of eukaryotic DNA replication is that the genome is precisely replicated once and only once each time a cell divides. To accomplish this remarkable feat in a eukaryotic system, replication initiates at several thousands, uniformly distributed chromosomal sites (Huberman and Riggs, 1966) called origins of DNA replication. Jacob, Brenner and Cuzin were the first to introduce the term 'origins' in their replicon model (Jacob et al., 1964). In this model (Fig. 1), specific sequence elements termed 'replicators' were postulated to genetically determine replication initiation sites on DNA molecules. The interaction of replicators with trans-acting regulatory factors, called initiators, activates initiation. Although proposed for *Escherichia coli*, this simple concept has been extremely useful in providing a framework to how the initiation of DNA replication occurs in all organisms.

Even though the replicon model has been a major driving force behind the discovery of eukaryotic initiator in unicellular yeast, it doesn't hold good for the multicellular eukaryotes, as novel regulatory mechanisms to spatially and temporally coordinate the initiation of DNA synthesis (Bell and Dutta, 2002; Kelly and Brown, 2000; Machida et al., 2005) have evolved in these organisms. One such mechanism is origin identification by eukaryotic replication initiator. As a result, even though highly conserved across various species, eukaryotic initiator's ability to determine origin sites varies greatly from species to species. With the increasing complexity of species, underlying mechanisms become more complex. Evidently, not even a single origin has been genetically characterized for the humans.

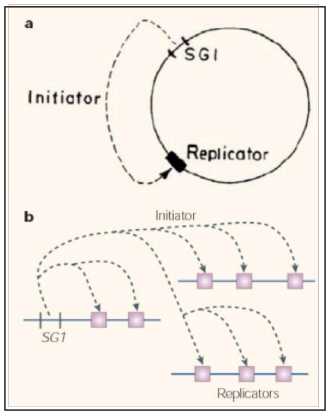

Fig. 1 The Replicon model a) Original concept of replicon model. The replicon is assumed to be a circular structure carrying two specific genetic determinants. A structural gene (SG1) determines the synthesis of a diffusible active element, the initiator. The initiator acts on a replicator, allowing the beginning of the replication (Jacob and Brenner, 1963) b) The presumed form of replicon model, applicable for eukaryotes with multiple origins (Gilbert, 2004).

Eukaryotic replicators

Identification of replicator sequences in single cell, budding yeast marks the first extension of replicon model to eukaryotic chromosomes. *S. cerevisiae* replicators were initially described as genomic sequences that confer on a plasmid the ability to be maintained extrachromosomally. These DNA elements were termed autonomously replicating sequences, or ARS elements (Stinchcomb et al., 1979). Purification of proteins that caused a specific DnaseI footprint on ARS1 led to the discovery of the eukaryotic initiator protein, ORC (Bell and Stillman, 1992). Molecular dissection of ARS1 (Fig. 2) revealed that it contains four essential elements for plasmid replication (Marahrens and Stillman, 1992). The most important element of ARS1 is the A element, which contains the ARS consensus sequence, or ACS. This 11 bp sequence, 5′(A/T)TTTA(T/C)(A/G)TTT(A/T)3′, is an essential feature of all known ARS elements (Newlon and Theis, 1993). The ACS constitutes half of the bipartite ORC binding site. ARS1 contains three additional elements, B1, B2, and B3 that together are also essential for origin function.

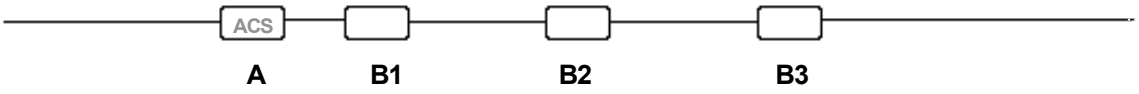

Fig 2 Schematic representation of ARS1. Boxes show four essential elements. The A element contains 11 bp, ARS consensus sequence (ACS). Although B1, B2 and B2 are not conseved within different ARSs, but essential for origin function.

Like *S. cerevisiae* ARS elements, *S. pombe* replicator sequences were identified through plasmid transformation studies (Clyne and Kelly, 1995). However, *S. pombe* ARS elements are much larger than their *S. cerevisiae* counterparts (0.5–1 kb versus 100–200 bp), and no well-defined consensus sequence analogous to the ACS has been identified that is essential for ARS function in *S. pombe*. Instead, *S. pombe* replicator elements are characterized by asymmetric stretches of adenine and thymine bases. Although in each ARS tested, deletion of small elements of 50 bp completely abrogates ARS function, these elements are not the same among different ARS elements, and they have no homology with one another aside from an unusually high A−T content (Kelly and Brown, 2000). Indeed, a computational genome wide analysis based on locating regions of DNA with higher A−T content than average identified 384 “A+T-rich islands” (Segurado et al., 2003). Twenty of these islands chosen at random were tested for ARS activity, and 18 were active origins, showing that A−T content is an excellent predictor of functional ARS elements. Genome wide screens for *S.pombe* and *S.cerevesiae* resulted in similar number of origins. Since these organisms have similar genome sizes and lengths of S phase, they contain similar numbers of replication origins.

The plasmid maintenance assay, which led to identification of replicator sequences in yeast, is not predictive for metazoans origins. Numerous attempts have been made to test the ability of many probable origins of replication in supporting the maintenance of episome. Any plasmid transfected into Drosophila

Schneider cells undergoes autonomous replication, regardless of the sequence contained on the plasmid (Smith and Calos, 1995). It is established now that in metazoans the efficiency of episome replication is more related to the length rather than the sequence of the insert (Marahrens and Stillman, 1992; Rowley et al., 1994; Theis and Newlon, 1994). Any DNA more than 15 kb long is maintained in this assay. Few origins have been characterized in metazoans using techniques, which enable mapping of physical location at which DNA synthesis initiates at a given locus. The three commonly used techniques are two dimensional gel electrophoresis, PCR analysis of nascent DNA strands and the replication initiation point mapping method (Cotterill, 1999). Despite the inability of investigators to isolate an autonomously replicating sequence in mammalian systems, approximately 20 mammalian origins (sites of replication initiation) have been identified (Todorovic et al., 1999). In some cases, DNA elements derived from these loci can direct replication at ectopic sites, and as such they are termed replicators. However, it is not entirely clear if the metazoan origins contain a single point of replication initiation or a zone with multiple initiation points.

A markedly less complicated example of mammalian replicator exists near the human lamin B2 gene. The lamin B2 origin was mapped to a 500 bp region 3′ of the lamin B2 gene by competitive PCR of nascent DNA strands (Giacca et al., 1994). Recently, *in vivo* crosslinking followed by chromatin immunoprecipitation demonstrated that the cell cycle dependent footprint of the lamin B2 origin is likely to be mediated by pre-RC components (Abdurashidova et al., 2003). However, in filter binding assays, HsORC was unable to distinguish between human lamin B2 and control sequences, except for its preference for A-T rich sites. So it seems likely that ORC does bind to and initiate replication from these regions, but it is not directed to these regions by binding to a specific DNA sequence.

Eukaryotic initiators

The first eukaryotic initiator protein was originally discovered in budding yeast through biochemical fractionation of nuclear extracts (Bell and Stillman, 1992). Aliquots of the extract were incorporated in DNase protection assays using ACS sequences of ARS1. These footprinting experiments determined that a six-protein complex, given the name origin recognition complex (ORC), specifically bound ARS1 in an ATP-dependent manner. In yeast, ORC is a heteromeric protein composed of six different subunits (Bell and Dutta, 2002). Each protein subunit is numbered according to decreasing size (Orc1-6), with Orc1 being the largest subunit at 104 kDa and Orc6 the smallest at 50 kDa. The precise structural mechanism of ORC binding to DNA is not clear.

It is evident that the ORC complex is conserved throughout eukaryotes. Soon after its discovery in *S. cerevisiae*, orthologs of each ORC subunit were identified in various other eukaryotic organisms, including *S. pombe* (Moon et al., 1999), *X. laevis* (Rowles et al., 1996), *D. melanogaster* (Gossen et al., 1995), *and H. sapiens* (Dhar and Dutta, 2000; Vashee et al., 2001). Identification of ORC was more difficult in other species as it was with budding yeast since individual ORC subunits are not as tightly associated in *S. pombe* and mammals. As a result, extraction of ORC in these species does not typically yield the full complex. Following characterization of different ORC complexes, it is clear that individual subunits share a great deal of similarity between species. The human homologs of Orc1, Orc2, Orc4, and Orc5 show considerable similarity to the yeast and metazoan equivalents. It is interesting to note that even though eukaryotes appear to have completely divergent replicators they all have a similar initiator binding to these different origin sites.

DNA binding by ORC

Owing to its well-defined origins, characterization of ORC association with DNA has been studied in more detail in budding yeast than in any other species.

Budding yeast ORC specifically binds to both the A and B1 elements of origins of replication in an ATP-dependent manner. A single base-pair mutation within either the A or B1 domains causes significant decrease in ORC association and in levels of DNA replication (Bell and Stillman, 1992; Diffley et al., 1994; Rowley et al., 1995). Throughout the cell cycle, budding yeast ORC is localized to ~460 chromosomal sites (Wyrick et al., 2001), spaced approximately every 20-30 kb. Interestingly, this number is significantly lower than the total number of ACS sequences identified throughout the *S.cerevisiae* genome. This suggests that there is some selectivity over origin sequences. In budding yeast, it is the five largest ORC subunits (ORC1-5) that are required to recognize and bind to DNA (Lee and Bell, 1997). ScOrc6p is the only subunit dispensable for origin DNA binding, though it remains essential for DNA replication and cell viability (Li, J. J. and Herskowitz, 1993). ScORC requires ATP with origin DNA. The Orc1, Orc4 and Orc5 proteins belong to the AAA^+ superfamily with a well-conserved ATP binding Walker A motif. In yeast, protein-ATP cross-linking studies have shown that in presence of origin DNA, Orc1 and Orc5 bind to ATP through Walker A motifs. ScORC with mutated Orc1 Walker A motif is unable to bind to the origin sequences. Mutation in Walker A motif of ScOrc5 does not affect ORC's DNA binding *in vitro,* but impairs the replication initiation ability *in vivo* (Klemm et al., 1997; Lee, D. G. et al., 2000).

In contrast to budding yeast, fission yeast exhibits a significantly different method of origin selection. Origins of replication contain similar AT-rich domains to budding yeast, yet ORC association with origins is dependent upon the AT-hook located on Orc4. It is this DNA binding motif found at the N terminus of the protein that mediates the DNA sequence specificity of ORC in fission yeast (Chuang and Kelly, 1999). Although not found in any other Orc4 homologs, the AT-hook in *S. pombe* Orc4 is essential for cell viability.

In metazoans, origins of replication and the nature of ORC binding are much less defined. As mentioned earlier, there does not appear to be a consensus

sequence linked to metazoan origins, and several factors are involved in ORC association with origin DNA. It has been difficult to study ORC association with origins, as it is clear that metazoan ORC possesses little selectivity over DNA sequences.

In Drosophila, ORC binds to DNA in an ATP dependent manner but shows no sequence specificity (Chesnokov, I. et al., 2001). However, it has been shown that complex has higher affinity for negatively supercoiled DNA (Remus et al.; Remus et al.). *In vivo* it is found to be associated specifically with chorion gene locus. Chorion genes are amplified in the follicle cells surrounding the developing oocyte. ORC mediated amplification of this locus provides an interesting system to study replication initiation *in vivo*. Like ScOrc1, DmOrc1 Walker A mutant do not interact with DNA even in the presence of ATP. Walker A mutants of Orc4 and Orc5 subunits do not affect the ability of DmORC to interact with DNA. Unlike ScOrc6p, DmOrc6p is required for DNA binding (Chesnokov, I. et al., 2001). For these analyses DmORC was purified from insect cell system. Similarly, HsORC was reconstituted by co-infection of baculovirus expressing individual ORC subunits in insect cells. Purified Human ORC has intrinsic DNA binding activity that is modestly stimulated by ATP (Vashee et al., 2003). Moreover, HsORC binds to and stimulates initiation from any DNA sequence using *in vitro* Xenopus replication system. Unlike DmORC, human ORC forms stable sub-complexes with Orc1 and Orc6 subunits in sub-stoichiometric levels. Although, this property proved useful in addressing ORC architecture (Ranjan and Gossen, 2006), the mode and specificity of ORC-DNA binding remains largely unclear. Walker A mutants of Orc1, Orc4 or Orc5 have shown to be defective in DNA replication *in vitro*. However, DNA binding by these mutants is reduced modestly in the same *in vitro* experiment (Giordano-Coltart et al., 2005).

In humans, it appears that not all of the ORC subunits remain chromatin-bound throughout the cell cycle. Human Orc1 is only found on chromatin during G1 phase and gets removed during DNA synthesis (Kreitz et al., 2001). This

translocation of Orc1 is thought to be a mechanism by which reformation of the pre-RC and re-replication of the genome is prevented within the same cell cycle. It is possible that in humans, Orc1 is required for the initial binding of ORC, but not for its maintenance at origins, thus it is dispensable for ORC function after it is loaded on to chromatin. It is clear that in all organisms examined thus far ORC is essential for DNA replication (Bell and Dutta, 2002; Kelly and Brown, 2000). As stated earlier, the primary function for ORC is the recognition of origins and subsequent loading of the pre- RC components.

In addition to its role in DNA replication, ORC has been found to function in several other cellular processes. For example, ORC is involved in the reorganization of chromatin. As part of this role, it appears as though ORC is involved in transcriptional silencing, which renders regions of chromosomes inactive, similar to heterochromatin in higher eukaryotes. The best-characterized example of ORC participation in gene silencing is in *S. cerevisiae* (Foss et al., 1993). It is known that ORC binds to regions adjacent to the silent mating type loci and recruits silent information regulator (Sir) proteins (Triolo and Sternglanz, 1996). The loading of Sir proteins results in a position effect, in which the genes around these loci are repressed. In humans, ORC has been found to interact with histone acetyltransferases (Burke et al., 2001). Acetylation of histones typically results in a loosening of the nucleosome, which generally allows genes to be transcribed. Although the significance of ORC interaction with HATs is unclear, it is possible that ORC could regulate the acetylation of adjacent areas ensuring that gene silencing is maintained.

Mechanism of origin specification

A situation where origin selection is completely random can be potentially fatal. Therefore, it is assumed that mechanisms evolved to select origins with some uniformity on chromosomes. If this would not be the case, there is a significant probability that large pieces of chromosomes would enter S phase with no

replication initiation protein assembly. This in turn would lead cells to mitosis with unreplicated DNA, at least if S phase check points would not be able to detect a weak replication activity in these chromosomal locations. Although studies show that metazoan ORC binds DNA with little or no sequence specificity, there has to be a mechanism to ensure the ORC's distribution on chromosomes in a way that enable timely replication of all chromosomes. There are strong indications that apart from DNA sequence of replicators, several other factors can play a pivotal role in origin specifications. Some of these are described here.

Transcriptional activity is probably the best characterized mechanism that plays a role in directing ORC to specific chromosomal sites. Strong evidence for a role of transcriptional activity in origin specification comes from Xenopus and Drosophila where replication initiates at random prior to the mid-blastula transition and becomes more specific concomitantly with the start of transcription after the MBT (Hyrein et al., 1995; Sasaki et al., 1999). Likewise, in budding yeast, nearly every origin of replication is found in an intergenic region, suggesting a mutual exclusivity between replication and transcription (Raghuraman et al., 2001; Wyrick et al., 2001). These studies suggest a role for transcription in negatively regulating ORC binding. However, examples of a positive interplay also exist. For example, transcription factor binding plays an important role in *S. cerevisiae* origin usage at ARS1, where binding of Abf1 facilitates plasmid maintenance of an ARS1 containing plasmid (Marahrens and Stillman, 1992; Marahrens and Stillman, 1994). Abf1 was found to regulate replication from this origin by limiting nucleosome binding within the origin (Lipford and Bell, 2001).

Next, methylation and acetylation appear to be another mechanisms that negatively regulate ORC binding. Methylation of plasmid DNA at CpG sites inhibits replication in Xenopus egg extracts, due to inhibition of ORC DNA binding (Harvey and Newport, 2003). Consistent with this observation, it has previously been shown that in mammalian cells, undermethylated regions of the genome often coincide with origins of replication (Delgado et al., 1998). Nascent

strand analysis revealed that the undermethylated regions are often associated with origins of replication, consistent with the notion that methylation is inhibitory for ORC binding, as found in Xenopus. In Drosophila, acetylated histones are localized to active origins at amplification foci, coincident with ORC (Aggarwal and Calvi, 2004). In this system, hyperacetylation of histone H4 leads to redistribution of ORC from amplification foci to a genome-wide staining pattern. Conversely, tethering of deacetylases to the DNA leads to a decrease in origin activity. In Xenopus eggs, injection of a plasmid containing a TATA-box and five GAL4-VP16 binding sites leads to localization of replication forks, with concurrent acetylation of histones in this transcriptional domain (Danis et al., 2004). In both of these systems, role of pre-RC proteins other than ORC in the interplay between origin activity and acetylation is not clear, but what is clear is that epigenetic events have a definite role in origin localization.

Another interesting possibility is that a more sequence specific DNA binding protein directs ORC to the specific DNA sites. For example, in Epstein Barr virus, EBNA1 protein binds with ORC and recruits it to viral replication origin, oriP. Moreover, mechanism of SpORC binding to DNA through N terminus A-T hook domain suggests the possibility of a more sequence specific protein directing ORC to origins of DNA replication. It should be noted however that if ORC is targeted to DNA by additional proteins, different sequence-specific proteins must be responsible at different foci. Otherwise, a unifying sequence would have emerged among known metazoan origins.

.

The pre-Replicative complex assembly

During G1 phase, a number of protein factors and complexes sequentially bind to origins of replication. The formation of the pre-replicative complex (pre-RC) is required to initiate DNA replication (Bell and Dutta, 2002; Kelly and Brown, 2000). At the heart of the pre-RC lies the origin recognition complex (ORC), which is responsible for locating and binding origins.

It is thought that ORC acts as a platform onto which other protein factors bind and initiate DNA synthesis. To commence the initiation process, two important loading factors bind to origin-associated ORC. The first protein to bind to ORC is the nucleotide-dependent loading factor Cdc6 (Perkins and Diffley, 1998). Once Cdc6 is bound to the initiation complex, Cdt1 recruits the mini-chromosome maintenance (MCM; MCM2-7) complex to the origin (Randell et al., 2006). Part, or the entire MCM complex acts as a helicase (Labib et al., 2000), assisting in the unwinding of the double helix. The assembly of ORC, Cdc6, Cdt1 and the six MCM proteins constitutes a pre-RC.

Once the pre-RC is formed, the origin is considered to be 'licensed', and thus competent for the initiation of DNA replication. As the cell cycle nears the end of G1, the levels of activated cyclin- and Dbf4-dependent kinases rise, and thus phosphorylation of Sld2 and Sld3 leads to the activation of the pre-RC (Zegerman and Diffley, 2007). These cyclin dependent kinases are involved in the recruitment of Cdc45 and the subsequent activation of the replicative helicases. As the MCMs begin unwinding the DNA, single-stranded DNA binding proteins (RPAs) stabilize the melted DNA strands and the polymerase-α/primase complex joins the assembly at origins to synthesize RNA primers. Once the DNA is replicated, pre-RCs are disassembled and disassociated from origins. To prevent re- replication of part or all of the genome within the same cell cycle, there are multiple, partially redundant mechanisms in place resulting from high Cdk activity (Li, C. J. et al., 2004). A consensus mode of pre-RC assembly is depicted in Figure 3.

As described before, pre-RC formation occurs in a sequential manner. ORC binding to DNA leads to Cdc6 and Cdt1 recruitment. Then the MCM complex is loaded to the origin and the complete pre-RC is formed.

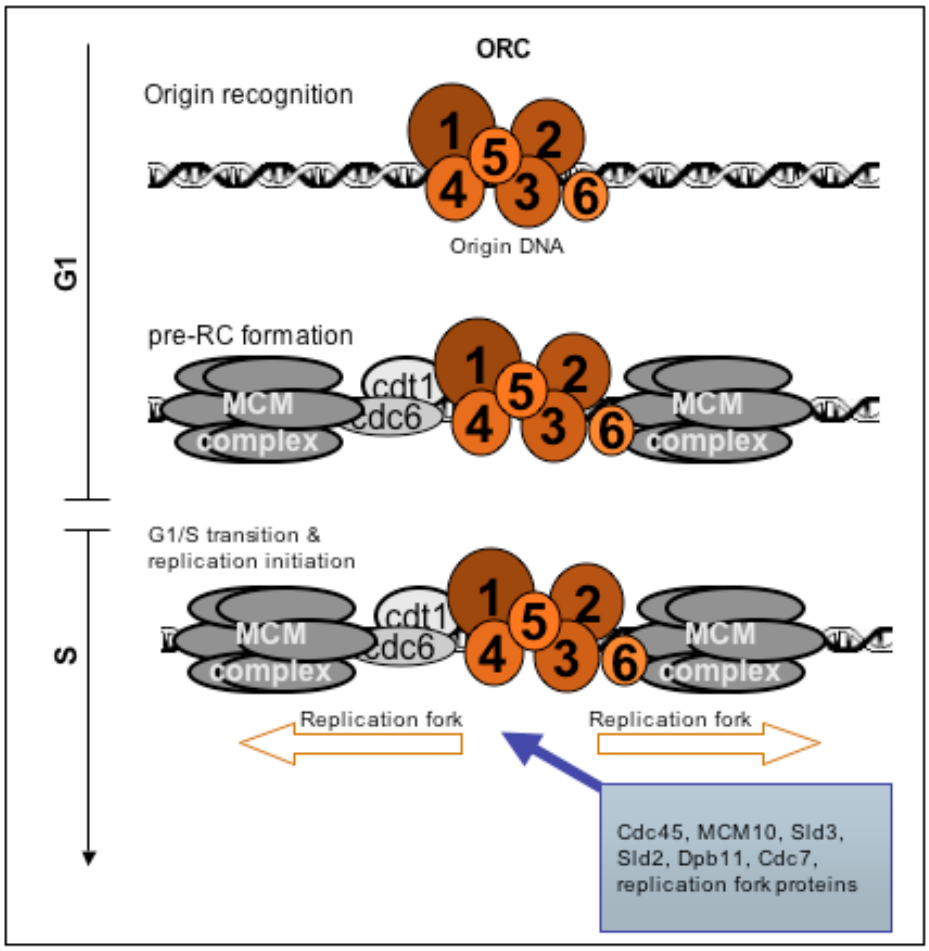

Fig. 3 A consensus model of eukaryotic pre-RC formation. ORC binds to the origin DNA in the late M phase and complete pre-RC (prereplicative complex) is formed during the G1 phase of cell cycle. Many other proteins associate with the chromatin bound pre-RC and begin bidirectional DNA synthesis at the start of S phase.

In metazoans, Geminin, an inhibitor of Cdt1 has been found to regulate the formation of pre-RCs as well. Geminin is degraded by the APC (anaphase promoting complex) during mitosis, and its very low levels during G1, allows the free Cdt1 to bind to origin and form the pre-RC. CDKs not only stimulate replication initiation but also prevent reinitiation. Inactivation of CDK is shown to induce re-replication in budding yeast.

Cdc6/Cdc18

Cdc6 protein (Cdc18 in *S. pombe*) is a member of the AAA+ ATPase family, as are several of the ORC subunits and the MCM proteins. The members of this family bind and hydrolyze ATP, and in doing so cause conformational changes to the protein. It is likely that this is the mechanism by which a controlled, sequential assembly of pre-RC components is performed (Neuwald et al., 1999). Orthologs of Cdc6 share a great deal of sequence and structural similarity (Crevel et al., 2005). The budding yeast protein sequence is approximately 25% identical and 45% similar to the *Xenopus, Drosophila* and human orthologs. Cdc6 is also similar to members of the ORC complex, particularly Orc1, in which it is 22% identical and 41% similar (Bell et al., 1995). Both mRNA transcript and protein levels for this protein peak at the M/G1 transition of the cell cycle. In G1 phase, Cdc6 initiates pre-RC assembly by binding ORC. Down regulation of Cdc6, at the mRNA or protein level prevents pre-RC formation in G1 and subsequent initiation of DNA replication. In budding yeast, Cdc6 is targeted for degradation following S phase (Piatti et al., 1995). In higher eukaryotes, Cdc6 appears to be stable throughout the cell cycle, but is actively transported from the nucleus after the start of S phase (Saha et al., 1998). However, there may be a small population of Cdc6 that remains localized to the nucleus throughout the cell cycle in mammalian cells (Alexandrow and Hamlin, 2004). The remaining Cdc6 found in the nucleus is thought to be involved in checkpoint surveillance. Cdc6 appears to have several roles leading up to DNA replication. Primarily, it is significantly involved in licensing origins of replication. Cdc6 is targeted to origins and the ORC complex during early G1 phase. In budding yeast, the ORC-Cdc6 complex has an enhanced DNA-binding specificity relative to ORC alone (Speck et al., 2005; Speck and Stillman, 2007). As mentioned earlier, it is thought that this increased specificity may be the key to metazoan origin identification. In budding yeast, Cdc6 binding to ORC causes ATP hydrolysis and a conformational change in ORC leading to the formation of an ORC-Cdc6 complex (Speck et al., 2005). This ORC-Cdc6 structure is thought to contain six

highly conserved AAA+ ATPase proteins, consisting of Cdc6 and five ORC (1-5) subunits. Reconstruction of electron microscopic images revealed that the ORC-Cdc6 complex forms a ring around the double helix (Speck et al., 2005; Speck and Stillman, 2007). Interestingly, along with several ORC subunits, Cdc6 contains regions conserved in clamp loaders, which function to load ring-shaped protein complexes onto DNA. Therefore, it is not surprising that Cdc6 is a key factor during pre-RC formation in early G1 phase (Cocker et al., 1996; Piatti et al., 1995), as it functions to load the MCM complex, which has been shown to form a ring structure with a central cavity the size of double-stranded DNA (Takahashi et al., 2004). Interestingly, inhibiting ATP hydrolysis by Cdc6 during pre-RC formation causes a stabilization of Cdt1 at origins (Randell et al., 2006), and as a result, MCM association with origins remains unstable. It is not until ATP hydrolysis of Cdc6 and the dissociation of Cdt1 that the MCM complex is tightly loaded at origins.

Cdc6 is also involved in cell cycle checkpoints, although this function is much less understood, as it appears to be species specific. In *S. cerevisiae*, Cdc6 plays a role in mitotic exit (Bueno and Russell, 1992), in *S. pombe* it is needed for the intra-S phase checkpoint (Murakami et al., 2002), whereas the *X. laevis* ortholog may play a role in monitoring both S and M phase progression (Clay-Farrace et al., 2003).

Cdt1 and Geminin

The replication licensing factor, Cdt1 (Cdc10-dependent transcript 1) gene was initially identified in *S.pombe* as a gene that is essential for DNA replication (Hofmann and Beach, 1994). Later studies in *Xenopus* and *S.pombe* established that Cdt1 associates with DNA and, like Cdc6/Cdc18, this association is dependent on the presence of ORC (Maiorano et al., 2000; Nishitani et al., 2000). MCM associations with DNA require the prior binding of Cdc6 and Cdt1 to the origin sites. It was shown for both humans and *Xenopus* that replication

dependent degradation of Cdt1 prevents re-replication in the same cell cycle and (Arias and Walter, 2004; Liu, E. et al., 2004) and this requires interaction of Cdt1 with PCNA through the consensus PIP box(Arias and Walter, 2006)

Geminin was discovered in a screening procedure designed to identify proteins that were degraded by mitotic *Xenopus laevis* egg extracts but not by interphase egg extracts (McGarry and Kirschner, 1998). Geminin is degraded by the APC at metaphase anaphase transition and it is known to inhibit the loading of MCM. Geminin binds to Cdt1 and inhibits its ability to load MCM proteins to the chromatin and form a pre-RC. The destruction of geminin at mitotic exit releases Cdt1, which can then serve to reload MCM protein on chromatin. This way geminin acts a negative regulator of origin licensing in metazoans during the S, G2 and M phases of cell cycle. Moreover, by physically interacting with transcription factors of the Hox and Polycomb families, geminin plays a major role during early eye and neuronal development (Del Bene et al., 2004; Luo et al., 2004). Geminin has turned up as an interesting candidate for studying the link between differentiation and proliferation, because of this dual role. Recent structural data of Geminin has revealed that it exists as homo-dimer and which often couples and form homo-tetramers (Okorokov et al., 2004; Thepaut et al., 2004).

Minichromosome maintenance (MCM) protein

The MCM complex is supposedly the final component of the pre-RC necessary to license origins of replication. It consists of six proteins, Mcm2 to Mcm7, with equal stoichiometry (Forsburg, 2004). Each MCM proteins were identified in yeast through genetic screens for mutations in plasmid maintenance. The MCM proteins are conserved members of the large AAA^+ family of ATPases and their homologs have been identified in all eukaryotes (Kearsey and Labib, 1998). MCMs are currently regarded as the prime candidates for the DNA helicases that unwind DNA at replication forks.

In contrast to the prokaryotic helicases where two helicases per origin are present, in average 10 to 40 MCM complexes bind to each replication origin in animal cells (Edwards et al., 2002). In yeast there seems to be more than 100 MCM complexes per replication origin (Lei et al., 1996). This successive loading occurs through a different mechanism, than the initial MCM complex. As mentioned earlier, this first MCM complex is stably loaded at origins following the ATP hydrolysis of Cdc6 and dissociation of Cdt1. Subsequent MCM complexes are loaded through a process that requires the continuous ATP hydrolysis of Orc1, and is dependent on the Orc1 and Orc4 proteins. The functional significance of this reiterative loading is unclear.

The MCM complex does not possess any intrinsic affinity for DNA (Mendez and Stillman, 2003), and thus relies on being loaded by pre-RC components onto origins. The precise mechanism by which MCM complexes unwind the DNA double helix is not known, but it is generally believed that the energy generated by ATP hydrolysis by the ATPase activity of the complex, is used in this process. During the S phase the MCMs are preferentially bound to the unreplicated DNA rather than the replicating or replicated DNA (Krude et al., 1996). Unlike RPA (Replication protein A) and PCNA, which are found at the replication fork during DNA replication, MCMs are not found at the site of new DNA synthesis during the S phase. Interesting models have been proposed to explain the role of MCMs as a helicase (Laskey and Madine, 2003; Schwacha and Bell, 2001), but there are some arguments against these models. As MCM holo-complex is devoid of any helicase activity *in vitro*. Only MCM4/6/7 complex have DNA helicase activity which, although weak, is sufficient to displace short oligonucleotides from complementary DNA circles (Ishimi, 1997; You et al., 1999). Recently, Cdc45/Mcm2-7/GINS complex, purified from protein extracts from Drosophila embryos, was shown to have ATP-dependent DNA Helicase activity *in vitro* (Moyer et al., 2006). It indicates that MCM association with other protein factors is required to form functional Helicase machinery.

Role of ORC in DNA replication: Exisiting Models

Apart from its central role in initiating DNA replication, ORC and its individual subunits have been implicated in diverse cellular activities in both the nucleus and the cytoplasm. ORC subunits were found at centrosomes, at the cell membranes, at the cytokinesis furrows of dividing cells, as well as at the kinetochore. Additionally, the function of ORCs in the establishment of transcriptionally repressed regions is described for many species and may be a conserved feature common for both unicellular eukaryotes and metazoans. However, many of the diverse functions of ORC are linked directly or indirectly to DNA replication control, especially origin selection or pre-RC formation.

There are two general models of ORC function in pre-RC assembly. One model suggests that ORC functions as a passive scaffold or “landing pad” to recruit other pre-RC proteins. In support of this model, ORC directly binds Cdc6 in vitro, and genetic studies suggest ORC interacts with Cdt1 and the Mcm2-7 complex (Bell and Dutta, 2002). An alternative model proposes that ORC actively participates in the assembly of other pre-RC components onto origin DNA (Davey et al.; Mendez and Stillman). Supporting the second model, a recent paper shows that hydrolysis of ATP by ScORC is required for successive MCM assembly at origins (Bowers et al., 2004). Again there are results in support of one model more than other and vice-versa. Thus, although ORC is required for pre-RC formation, its mechanism of action during this process is yet to be understood.

Single Chain Tetracycline Transregulators

Tet repressor is one of the best-characterized, sequence-specific DNA binding proteins. It is derived from the Tetracyline resistance element of Tn10 operon in *E.coli*. In absence of tetracycline, TetR binds to target sequences efficiently. However, introduction of inducer results in conformational change in protein and the binding is disrupted rapidly. The efficient DNA binding and rapid inducer

response provides a tightly controlled on/off mechanism. This system is used extensively for the eukaryotic gene-regulation. While unmodified TetR acts as a transcriptional repressor in plants and lower eukaryotes, it can be, but not always is efficient in mammalian cells. A consistently functional version for yeasts, flies and mammalian cell lines is TetR fused to a eukaryotic regulatory domain. Here, virion protein 16 from Herpes simplex (VP16) is fused to the C-terminal of TetR to make tTA, tetracycline trans-activator (Gossen and Bujard, 1992).

In general, prokaryotic regulatory proteins have been found to bind to DNA sequences that exhibit dyad symmetry (Pabo and Sauer, 1984). Since these proteins are usually dimers or tetramers, this suggests that the interaction with each operator half-side is identical. The palindromic DNA sequences recognized by TetR are called tet operators or tetO. There are two different sequences O1 and O2. The two wild type operators O1 and O2 share a nearly identical nucleotide sequence except at the central, seventh, and ninth base pairs (Fig 4a). The binding efficiency of TetR to these sequences differs only by a factor of 2. The crystal structure of TetR- tetO (Orth et al., 2000) reveals the amino acid mutations sufficient to switch the recognition specificity from the wild type tetO.

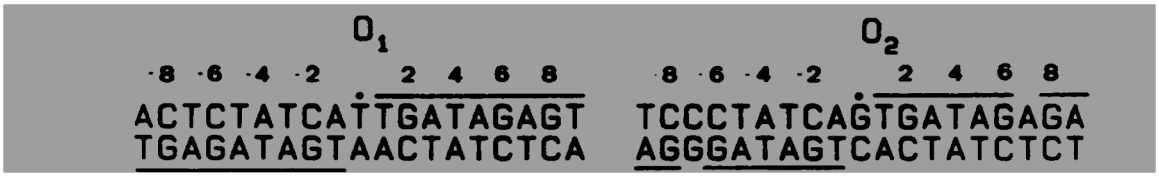

Fig. 4a tet operator sequences. Nucleotide sequences of wild type tet operators O1 and O2 (Wissmann et al., 1986).

TetR forms a homo-dimer with size of each monomer of about 25kD. Aided with the crystal structure information, dimerization domains have been identified. The N terminus binds to DNA and C terminus forms the regulatory domain. Upon binding with doxycycline TetR undergoes complex conformational changes. This results in the rapid disruption of DNA binding. TetR is the basis of wide variety of trans regulator based doxycycline controlled expression systems in eukaryotes.

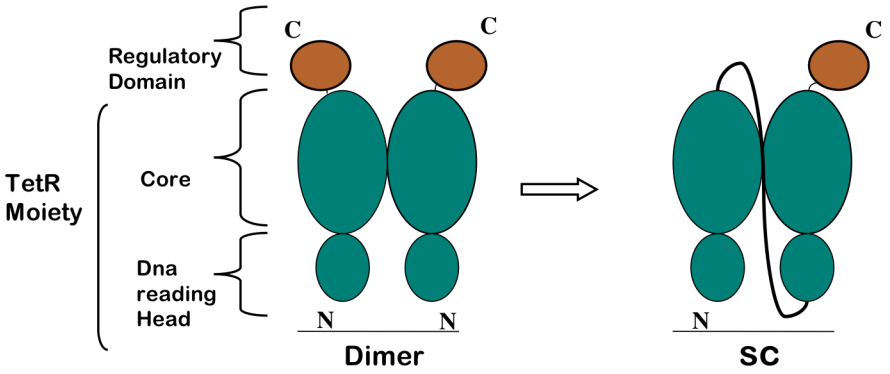

Fig. 4b Schematic representation of Tet repressor dimer and single chain version of Tet repressor. N terminal of Tet repressor (on the left) binds to DNA and C terminal is the regulatory domain. A flexible linker of 25 $(SG_4)_5$ amino acid is used to construct the single chain tet repressor (on the right)

In a recent work, C terminal of the first monomer is fused to the N terminal of second monomer (Fig. 4b) by a 25 $(SG_4)_5$ amino acid long flexible linker to generate a single chain Tet trans regulators (Krueger et al., 2003). To avoid the genetic instability due to identical sequence repeats all single chain regulators contain one wild type and one synthetic tetR gene. The monomerization has not compromised with the specificity and sensitivity of the tet trans regulators. With only one regulatory domain, single chain trans regulators will minimize the steric hindrances and potential influence on the activity of trans regulator as well as the fused domain.

Aim of the study

The architecture of the metazoan origins is poorly understood. In particular, no consensus DNA binding sites have been identified, either by genetic or biochemical means. This poses a serious bottleneck in the elucidation of the underlying mechanism of DNA replication initiation. It is now widely accepted that factors other than DNA sequences are responsible for origin identification. One of the opinions is that a more sequence specific protein directs ORC to the origins. Moreover, an existing model argues that ORC merely acts as a scaffold in the pre-RC assembly. If both the arguments are true, we hypothesized that recruitment of ORC to specific DNA sites could be sufficient to initiate pre-RC assembly, which in turn would trigger DNA replication, preferentially at these sites. The main aim of this study is to design a system where engineered human ORC can be recruited artificially to defined DNA sites and triggers replication activity at these sites. This would provide an artificial but useful system to study mechanisms that cannot be addressed to date, owing to the absence of any genetically characterized origin. Intrinsic DNA binding activity, which is a unique characteristic of human ORC can be potentially addressed with such a tool. I decided to use a single chain version of Tet repressor, fusing it to the N-terminus of various ORC subunits.