Review

Nuclear DNA replication initiation in kinetoplastid parasites: new insights into an ancient process

Calvin Tiengwe[1,2], Catarina A. Marques[1], and Richard McCulloch[1]

[1] The University of Glasgow, Wellcome Trust Centre for Molecular Parasitology and Institute of Infection, Immunity and Inflammation, Sir Graeme Davis Building, 120 University Place, Glasgow, G12 8TA, UK
[2] The Johns Hopkins University School of Medicine, Department of Cell Biology, Baltimore, MD, USA

Nuclear DNA replication is, arguably, the central cellular process in eukaryotes, because it drives propagation of life and intersects with many other genome reactions. Perhaps surprisingly, our understanding of nuclear DNA replication in kinetoplastids was limited until a clutch of studies emerged recently, revealing new insight into both the machinery and genome-wide coordination of the reaction. Here, we discuss how these studies suggest that the earliest acting components of the kinetoplastid nuclear DNA replication machinery – the factors that demarcate sites of the replication initiation, termed origins – are diverged from model eukaryotes. In addition, we discuss how origin usage and replication dynamics relate to the highly unusual organisation of transcription in the genome of *Trypanosoma brucei*.

Genome replication: a central process in all organisms

One facet of life is common to every organism, irrespective of whether it is a virus, prokaryote, or single- or multicelled eukaryote: transmission of the genome. Genome transmission involves copying the genetic material and passing it into offspring, thus propagating life. For double-stranded DNA genomes, replication using base complementarity drives transmission, allowing the generation of genome copies that are then partitioned during cell division. Although DNA replication is conceptually simple and mechanistically conserved between the three kingdoms of life, the machinery and regulation involved are highly complex and display perhaps surprising levels of variation between kingdoms [1,2]. As we will describe here, recent work has begun to examine the machinery and coordination of nuclear DNA replication in kinetoplastid protozoan parasites, agents of debilitating and often lethal human and animal diseases, and has revealed both conservation and divergence in this key cellular process relative to other eukaryotes.

Corresponding author: McCulloch, R. (richard.mcculloch@glasgow.ac.uk).
Keywords: kinetoplastid; *Trypanosoma*; DNA replication; origin recognition complex; origin of replication.

1471-4922/$ – see front matter
 http://dx.doi.org/10.1016/j.pt.2013.10.009

Understanding DNA replication is important not merely for mechanistic clarification. The machinery and process of DNA replication intersects with and influences many aspects of cell function, including gene expression, mutation, recombination, and chromosome segregation. Such issues pertain to survival of all cells, but may have particular resonance for pathogens. For instance, the life cycle of *Trypanosoma brucei* involves several stages at which differentiation occurs between replicative and non-replicative forms, allowing establishment within and transmission between the mammal host and tsetse fly vector. Great progress has been made in understanding the signals for these events and the gene expression changes that occur [3], but how DNA replication is activated or suppressed is

Glossary

AAA+ family proteins: ATPases associated with a variety of activities; enzymes with a conserved P-loop NTPase domain that contribute to many cellular functions, including nucleic acid biology, protease digestion, and acting as chaperones.
CMG complex: a higher order structure composed of CDC45, the MCM helicase complex, and the GINS complex. The MCM helicase acts to unwind DNA during replication, but displays optimal activity only once it is associated with CDC45 and GINS.
GINS complex: a heterotetramer composed of the proteins Psf1, Psf2, Psf3, and Sld5, which travels with MCM and CDC45 during replication as the CMG complex.
Minichromosome maintenance complex (MCM): the replicative helicase responsible for unwinding the DNA helix in eukaryotic replication; composed of six subunits, MCM2–7, which form a two-tiered hexameric ring. Interacts with Cdt1.
Origin recognition complex (ORC): the multi-subunit complex found in eukaryotes that binds to origins. The ORC is normally considered to be composed of six subunits, Orc1–6. Interacts with Cdc6, which is evolutionarily related to the Orc1 subunit.
Pre-initiation (pre-IC) complex: an activated form of the pre-RC, which is thought to be formed in S phase due to the recruitment of further factors, including MCM10 and the CMG complex, allowing recruitment of the replicative DNA polymerases.
Pre-replication complex (pre-RC): a higher order of structure of ORC–Cdc6–Cdt1–MCM, which is thought to form on origins prior to S phase.
Replication origin: locus in the genome where DNA synthesis begins during replication. In eukaryotes, the multiple origins used are not normally defined by conserved sequences; the exception to this is in *Saccharomyces cerevisiae*, where origins are composed of related autonomous replicating sequences (ARSs).
Walker A and B motifs: distinct, conserved sequence motifs found in ATP binding proteins; first reported in 1982 by J.E. Walker and colleagues.
Winged helix (WH) domain: a variant of the highly conserved helix-turn-helix domain, found in prokaryotic and eukaryotic proteins acting in both replication and transcription; thought to mediate protein–DNA and protein–protein interactions.

unknown. DNA replication normally generates high fidelity genome copies. However, some genome plasticity is crucial to provide adaptability, and this appears central to pathogen survival strategies, which can result in spectacular levels of variation. Several species of *Leishmania*, for example, display a remarkable capacity for gene copy number variation, which is seen as amplification of chromosome segments and ploidy changes of whole chromosomes, at least some of which is revealed by drug selection [4]. *T. brucei*, in common with many pathogens, survives in the face of host immune attack by antigenic variation, the periodic switching of surface antigens. In *T. brucei*, antigenic variation involves switches in variant surface glycoproteins (VSGs), a reaction that involves high rates of recombination among a huge *VSG* gene archive in the chromosome telomeres and subtelomeres [5]. If and how DNA replication might influence these and other examples of genome plasticity has been little studied.

Nuclear DNA replication in eukaryotes

In all organisms, DNA replication involves the ordered recruitment and activation of multiple proteins to genome sites at which DNA synthesis begins, termed origins (see Glossary). In eukaryotes, this cascade involves several protein complexes, with mediators acting to foster interaction between them (Figure 1). In this section, we provide an overview of replication initiation and refer the reader to several excellent reviews for more detail. The earliest acting factor in the replication cascade is termed the origin recognition complex (ORC), which is composed of six subunits (Orc1–Orc6) and binds the origin. Unlike in bacterial and some archaeal genomes, where conserved sequences define a single origin, the linear chromosomes of eukaryotes examined to date are replicated from multiple origins [1,2]. Moreover, in most eukaryotes origins are not conserved sequences but appear to be defined by DNA structure and chromatin environment [6,7]. The ORC subunits 1–5 are AAA+ family proteins, whereas Orc6 is unrelated [8]. The ORC forms a ring-shaped structure whose conformation is altered by binding to Cdc6 [9], another AAA+ protein closely related to Orc1 [8]. Such conformation changes are driven by ATPase activities of the subunits and allow the ORC to provide a platform to recruit further replication factors.

Eukaryotic DNA replication normally occurs once per cell cycle, and replication factor recruitment and activation dictates this. During the mitosis to G1 transition, the ORC binds to all potential origins in the genome. Cdc6 binds to Orc1 during G1 [6,7], stabilising the ORC on origin DNA [9] and allowing recruitment of the replicative helicase, a heterohexamer termed the minichromosome maintenance complex (MCM) whose subunits (MCM2–7) are also AAA+ proteins [10]. A further factor is critical for MCM loading: Cdt1 binds MCM and mediates interaction with the ORC, most probably through Orc6 [11]. The MCM complex is loaded as an inactive double hexamer and Cdt1 is released, completing the assembly of the pre-replicative complex (pre-RC), which renders the origins licensed for replication. At the onset of S phase, Cdc45 and the GINS complex (Sld5, Psf1, Psf2, and Psf3) are recruited and associate with MCM, forming the active helicase (CMG complex) [12].

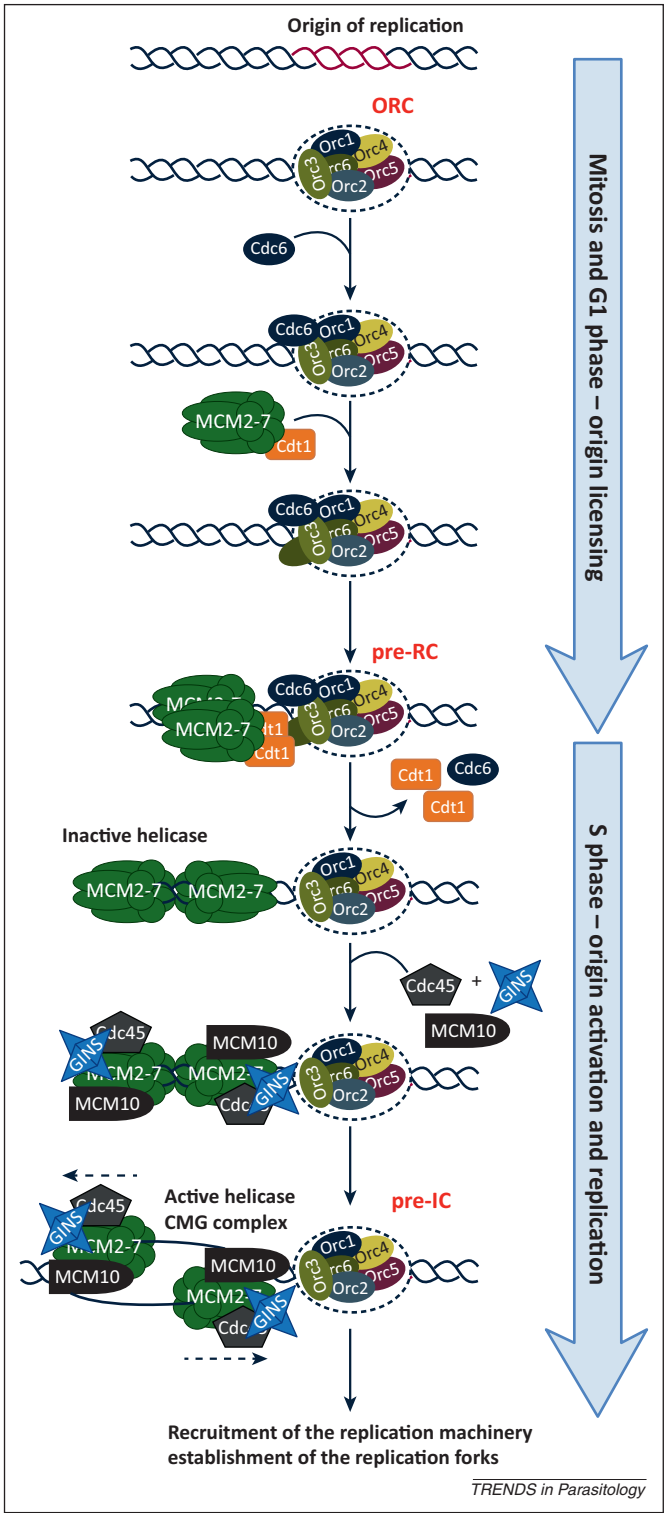

Figure 1. Cascade of factor recruitment during replication initiation in eukaryotes characterised to date. A diagram is shown detailing the order in which replication factors are recruited to origins of replication and the timing of these events during the cell cycle. Origins are first bound by the origin recognition complex (ORC), which is composed of six subunits (Orc1–6); relative locations of the subunits are based on electron microscopy analysis of the yeast complex [9]. Cdc6 binds and modifies the ORC to allow recruitment of the minichromosome maintenance (MCM) helicase, which is also a hexamer (MCM subunits 2–7); this recruitment step is mediated by MCM-bound Cdt1 and forms the pre-replication complex (pre-RC), licencing origins prior to S phase. Upon entering S phase, Cdc6 and Cdt1 are removed from the pre-RC (in some cases degraded or excluded from the nucleus; not shown), allowing recruitment of Cdc45, the GINS complex (Sld5, Psf1, Psf2, and Psf3; see Table 1 and Figure 3) and MCM10, forming a pre-initiation complex (pre-IC). The DNA helix is then opened and further replication factors recruited (not shown). Regulatory processes that modify the various complexes during the cell cycle, for instance, to prevent re-replication, are not shown.