Replicon hypothesis revisited

Hisao Masai

Department of Basic Medical Sciences, Tokyo Metropolitan Institute of Medical Sciences, 2-1-6 Kamikitazawa, Setagaya-ku, Tokyo 165-8506, Japan
Tel: +81-3-5316-3220
E-mail: masai-hs@igakuken.or.jp

Abstract

Nearly 70 years after the proposal of semiconservative replication of generic material by Watson and Crick, we now understand many of the proteins involved in the replication of host chromosomes and how they operate. The initiator and replicator, proposed in the replicon hypothesis, are now well defined in both prokaryotes and eukaryotes. On the other hand, studies in prokaryotes and Archae indicate alternative modes of initiation, which may not depend on an initiator. Here I summarize recent progress in the field of DNA replication and discuss the evolution of replication systems.

The replicon hypothesis proposed by Jacob, Brenner and Cuzin 50 years ago (1), has played a leading role in studies of the regulation of DNA replication. It proposes an initiator and replicator as components of a replicon, the unit of DNA replication. The hypothesis was genetically substantiated by the discovery of *dnaA* by Kohiyama (2) and others, and of *oriC*, by Yasuda and Hirota (3) and others. The search for counterparts in eukaryotes led to the discovery of ARSs as candidate chromosome replicators (4), and subsequent studies demonstrated that these sequences serve as bona fide replicators of eukaryotic chromosomes. Initiators for eukaryotic chromosome replication had remained elusive until Bell and Stillman discovered ORC in budding yeast (5). Orthologs of ORC were soon identified in other species including humans, supporting the idea that ORC is the initiator for eukaryotic chromosome replication. These findings suggested the universality of replicon-based replication throughout evolution. In this article, I will first briefly summarize recent progress in this field and then discuss the

replicon hypothesis in view of the evolution of replication systems.

Initiation of DNA replication

In bacteria, the initiator DnaA protein binds to the replicator, *oriC*, and positively regulates initiation. This causes melting of the DUE (DNA unwinding element) duplex DNA near the initiator binding site, leading to the loading of replicative helicases (6). Duplex opening is facilitated by sharp bending of origin DNA by IHF and single-stranded DNA binding activity of DnaA at the DUE. ATP-ADP conversion of DnaA regulates initiation. RIDA (Regulatory inactivation of DnaA), DDAH (*datA*-dependent DnaA-ATP hydrolysis), and SeqA binding to the newly synthesized *oriC* DNA inhibit re-replication, whereas DARS (DnaA-reactivating sequences) dependent regeneration of the ATP-bound form of DnaA promotes timely initiation (7).

In eukaryotes, replication proceeds in two-stages (8). During G1, pre-RC is assembled at prospective initiation sites on the chromosomes. This process, called licensing, is strictly regulated by many layers of mechanisms so that it occurs only during G1. This is essential to ensure that replication of the genome occurs once and only once during a cell cycle. Failure will severely increase genome instability, leading to cancer formation. In the second stage, the pre-RC is activated by phosphorylation reactions, catalyzed by two kinases, Cdc7 and CDK, at the transition from G1 to S phase. This leads to assembly of an active replicative helicase, CMG (Cdc45-MCM-GINS), engaging three DNA polymerases (see below).

Budding yeast ORC binds specifically to the replicator ARS (5), although the specificity of ORC binding in higher eukaryotes is still not very clear. ORC binding determines potential origins and Mcm is recruited with the help of Cdc6 and Cdt1. Pre-RC assembly and activation of replication initiation has been reconstituted in yeast with purified proteins (9). The active replication fork has been reconstituted with human factors as well (10). The reconstitution of replicator-dependent initiation on a chromatin template with purified proteins has shown that binding of CMG helicase at the origin induces splitting of a double hexamer, DNA untwisting and 3 bp melting within each Mcm ring, which is promoted and stabilized by the MCM6 wedge structure (11).

DNA chain elongation

Following initiation at origins, replicative DNA helicases are engaged at replication

forks to promote DNA synthesis mediated by DNA polymerases. In *E. coli*, a single polymerase, DNA polymerase III holoenzyme, drives the replication fork on both strands (12). In eukaryotes, different DNA polymerases are required for leading and lagging strand DNA synthesis; DNA polymerase ε for leading and DNA polymerases α and δ for lagging strands. It is proposed that both strands are concurrently replicated at the fork by "snap back" of the lagging strand template DNA.

PCNA is an evolutionally conserved DNA polymerase associated factor, which functions as a clamp and facilitates processive DNA synthesis. It also recruits a number of replication/ repair factors to maintain the integrity of the replication fork complex and enable it to swiftly respond to various unscheduled events to repair damaged templates and resume DNA synthesis. PCNA and its loader, RFC, are well conserved from bacteria to humans (12).

Cellular responses to stalled replication forks.

Once cells commit to initiation, DNA replication needs to proceed until the entire genome is replicated. However, replication fork progression can be stalled or halted by a number of external and internal impediments. Cells need to cope with these situations, first, by detecting the stalled fork, and second, by restoring an active fork (13). In bacteria, a conserved DEXH-type helicase, PriA, plays a central role in both steps (14). In eukaryotes, the replication fork complex (replisome) is significantly more complex than its prokaryotic counterpart, partly because it needs to deal with templates wrapped in chromatin, and also because tens of thousands of simultaneously progressing replication forks need to be coordinated. In fact, eukaryotic replisomes contain factors involved in chromatin remodeling (such as FACT) and replication stress checkpoint (Claspin, Tim-Tipin, And-1 *etc*.; 12). The latter factors are essential for both replication stress checkpoint induction and efficient replication fork progression (15).

Replication timing regulation

On human chromosomes, replication is initiated at 20,000~30,000 replication origins, and replication occurs in a spatially and temporally regulated manner (16,17). Early firing origins cluster in early-replicating domains and form minute foci spread through nuclei, while, during mid-to-late S, replication foci are located at the nuclear periphery and around nucleoli. Near the end of S phase, replication occurs mostly at large foci in

heterochromatic regions. Replication timing is affected by multiple factors including the epigenomic state of chromatin. A model was proposed that the stoichiometry of Mcm loading is correlated with the firing timing of origins during S phase (18). Other factors specifically regulating replication timing have also been reported. Rif1 affects genome-wide replication timing by suppressing mid-S firing origins using two mechanisms (19, 20, 21). First it recruits PPase1 to counteract phosphorylation required for origin firing (22,23,24) and second, it generates chromatin architecture that is closely related to mid-S replicating domains (25,26). Polθ was reported to regulate replication timing by regulating Mcm loading at some late-firing origins (27). Some late replicating domains, that are not regulated by Rif1, are associated with repressive histone marks including HH3 K9 methylation (28). Thus, temporal and spatial regulation of genome replication in S phase is regulated by multiple factors in different chromatin compartments defined by Rif1, epigenomic states, and pre-RC formation.

Diversity in modes of genome replication: an alternative mode of *E.coli* genome replication

Bacterial genomes are replicated efficiently by the actions of the DnaA initiator and replicator *oriC* (6). Because of this efficiency, this system is suitable for competition with other bacteria and is common to all bacterial replicons. However, this system consumes high amounts of energy with protein synthesis being required in each cycle of replication for *de novo* synthesis of the initiator. Thus, it is labile in nutrient-limiting environments, and *E.coli* has developed an alternative pathway for replicating its genome in a manner independent of DnaA/*oriC* (29). This second mode of replication, coined SDR (stable DNA replication), occurs in the absence of ongoing protein synthesis. The most robust SDR is observed in *rnhA* mutants, lacking cellular RNaseH activity. *E.coli* can grow in the absence of *dnaA* or *oriC* in *rnhA* mutants, likely due to RNA-DNA hybrids/R-loops generated on the genome. SDR was initially discovered in thymine-starved cells where *oriC*-dependent DNA synthesis was arrested. SDR-like replication also occurs in cells entering stationary phase, or in certain genetic backgrounds such as *recG* or *dam* methylase mutants (30). This mode of replication may be induced upon sudden changes in the environment or upon encountering cellular stresses that challenge the course of DNA replication.

In comparison with the dnaA-*oriC* mediated initiation, which occurs with

100% efficiency and in a highly sequence-specific manner, SDR is inefficient and is characterized by low sequence specificity. However, because it is energy-saving, and suitable for survival during nutritionally challenging conditions, it can be more robust than the DnaA-*oriC* replication in certain situations. This mode of initiation is not suitable for competition through other bacteria, but may be useful during adaptation to changing environments. It could increase the survival of the species by elevating the mutation rate during replication.

Alternative modes of replication are not restricted to *E.coli*. In archae, Haloferax volcanii, deletion of all four replication origins did not affect growth, and replication initiated promiscuously (31). The origin-less strain depends on RadA, the RecA homologue of this species, for growth. In cyanobacteria, strains lacking DnaA were identified, suggesting the existence of a DnaA-independent pathway for DNA replication. In viral DNA replication, it has been well documented that genomes can be replicated by multiple mechanisms, including mechanisms dependent on recombination. In T4 phage DNA replication, the genome can be replicated from origins that generate RNA-DNA hybrids as well as in the form of RDR (Recombination-Dependent Replication; 32).

A primordial mode of DNA replication independent of initiator?

The replicon hypothesis explains the regulation of replication in bacteria very well. Orc-ARS in yeast also fits this hypothesis. However, the initiation of DNA replication appears to be more plastic and adaptive, and potentially more stochastic in eukaryotes (33).

Although specific sequences are clearly critical for replication origins in yeast, identifying similar type sequences in higher eukaryotes has been difficult. ARS sequences have not been identified in mammalian cells. However, a number of methods have been developed to identify cellular replication origins (34,35,36,37,38), and we now can clearly identify where replication will start in given cell types. Conserved sequences have not been identified, but more than 50% of mapped origins are associated with potential G4 forming sequences, raising the possibility that G4 may play a role in origin regulation (39,40,41).

The alternative mode of bacterial replication requires transcription and potentially the formation of RNA-DNA hybrids/R-loop, and could initiate at multiple

locations on the genome, reminiscent of initiation in eukaryotes. This mode of replication is simple, since it could operate without a sequence-specific initiator as long as a replicative helicase can be loaded onto duplex DNA.

Some replication factors, such as Clamp and Clamp loader, are highly conserved between prokaryotes and eukaryotes. On the other hand, some other core replication factors including helicases, DNA polymerases, and primase are more structurally divergent, suggesting that they are derived from separate ancestral proteins (12). Initiators (DnaA and Orc) are AAA^+ proteins, but their modes of action could be distinct, and they may also have evolved separately.

I speculate that the primordial mode of DNA replication would not require initiators and replicators, but would be initiated at some structured nucleic acid where a replicative helicase can be loaded. Alternatively, replication may have occurred in the absence of active helicases, facilitated by RNA-driven catalysis. In the alternative mode of *E. coli* chromosomal replication, replication can start wherever transcription occurs and R-loops are formed. RNA-DNA hybrid/ R-loop formation is greatly facilitated by the presence of a G-rich sequence capable of forming G-quadruplex structures (42), thus making the R-loop/G4 structure an attractive candidate for a prototype of an ancient mode of replication.

The prevalent DnaA-*oriC* replication, that dominates the bacterial kingdom, may have evolved because of a need for efficient replication by horizontal transfer of a paired initiator-target unit. Thus, current bacteria are equipped with highly evolved replication systems, that specialize in replicating the 4.6 Mb genome in 40 min every 20 min. Eukaryotes may have acquired ORC-replicator units independently. The prototype initiator-replicator interactions were probably sequence-specific, as still observed in some yeasts. However, as species evolved, they maintain the initiator (ORC)-replicator interactions, but the sequence specificity was relaxed to increase the flexibility and robustness required to replicate the large genomes. The sequence- and initiator-independent replication modes could be conserved in eukaryotes where it may function in specific cell types or under some stressed conditions. Search for such replication systems in eukaryotes is now ongoing.

Acknowledgments

I truly congratulate Ernesto for his 90th birthday and for his great contributions to science and to BBRC. I wish Ernesto continued good health and success. I would like to thank Dr. Taku Tanaka in my laboratory for his work on *E.coli* replication and for useful discussions. I also thank Dr. Jun Horiuch for correcting English.

References

1. Jacob, F., Brenner, C. & Cuzin, D. Cold Spring Harbor Quant. Biol. 28 (1963), 329–348.

2. Emmerson PT, Kohiyama M. Effect of the mutations recB and polA on gamma-ray-induced DNA synthesis in a temperature sensitive dnaA mutant of Escherichia coli K12. Nature. 233 (1971), 214-215.

3. Yasuda S, Hirota Y. Cloning and mapping of the replication origin of Escherichia coli. Proc Natl Acad Sci U S A. 74 (1977), 5458-5462.

4. Struhl K, Stinchcomb DT, Scherer S, Davis RW. High-frequency transformation of yeast: autonomous replication of hybrid DNA molecules. Proc Natl Acad Sci U S A. 76 (1978), 1035-1039.

5. Bell SP, Stillman B. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. Nature. 357 (1992), 128-134.

6. Katayama T. Initiation of DNA Replication at the Chromosomal Origin of E. coli, oriC. Adv Exp Med Biol 1042 (2017), 79-98.

7. Sakiyama Y, Nagata M, Yoshida R, Kasho K, Ozaki S, Katayama T. Concerted actions of DnaA complexes with DNA-unwinding sequences within and flanking

replication origin oriC promote DnaB helicase loading. J Biol Chem. 298 (2022), 102051.

8. Diffley JF. Regulation of early events in chromosome replication. Curr Biol. 14 (2004), R778-786.

9. Yeeles JT, Deegan TD, Janska A, Early A, Diffley JF. Regulated eukaryotic DNA replication origin firing with purified proteins. Nature 519 (2015, 431-435.

10. Baris Y, Taylor MRG, Aria V, Yeeles JTP. Fast and efficient DNA replication with purified human proteins. Nature 606 (2022), 204-210.

11. Lewis JS, Gross MH, Sousa J, Henrikus SS, Greiwe JF, Nans A, Diffley JFX, Costa A. Mechanism of replication origin melting nucleated by CMG helicase assembly. Nature 606 (2022), 1007-1014.

12. Yao NY, O'Donnell ME. The DNA Replication Machine: Structure and Dynamic Function. Subcell Biochem. 96 (2021), 233-258.

13. Masai H. TT-pocket/HIRAN: binding to 3'-terminus of DNA for recognition and processing of stalled replication forks. J Biochem 172(2022), 57-60.

14. Masai H, Tanaka T, Kohda D. Stalled replication forks: making ends meet for recognition and stabilization. Bioessays. 32 (2010), 687-697.

15. Yeeles JTP, Janska A, Early A, Diffley JFX. How the Eukaryotic Replisome Achieves Rapid and Efficient DNA Replication. Mol Cell. 65 (2017), 105-116.

16. Zhao PA, Rivera-Mulia JC, Gilbert DM. Replication Domains: Genome Compartmentalization into Functional Replication Units. Adv Exp Med Biol 1042 (2017), 229-257.

17. Masai H, Matsumoto S, You Z, Yoshizawa-Sugata N, Oda M. Eukaryotic chromosome DNA replication: where, when, and how? Annu Rev Biochem. 79 (2010), 89-130.

18. Dukaj L, Rhind N. The capacity of origins to load MCM establishes replication timing patterns. PLoS Genet 17 (2021), e1009467.

19. Hayano M, Kanoh Y, Matsumoto S, Renard-Guillet C, Shirahige K, Masai H. Rif1 is a global regulator of timing of replication origin firing in fission yeast. Genes Dev. 26 (2012), 137-150.

20. Yamazaki S, Ishii A, Kanoh Y, Oda M, Nishito Y, Masai H. Rif1 regulates the replication timing domains on the human genome. EMBO J. 31 (2012), 3667-3677.

21. Cornacchia D, Dileep V, Quivy JP, Foti R, Tili F, Santarella-Mellwig R, Antony C, Almouzni G, Gilbert DM, Buonomo SB. Mouse Rif1 is a key regulator of the replication-timing programme in mammalian cells. EMBO J. 31 (2012), 3678-3890.

22. Hiraga S, Alvino GM, Chang F, Lian HY, Sridhar A, Kubota T, Brewer BJ, Weinreich M, Raghuraman MK, Donaldson AD. Rif1 controls DNA replication by directing Protein Phosphatase 1 to reverse Cdc7-mediated phosphorylation of the MCM complex. Genes Dev. 28 (2014), 372-383.

23. Davé A, Cooley C, Garg M, Bianchi A. Protein phosphatase 1 recruitment by Rif1 regulates DNA replication origin firing by counteracting DDK activity. Cell Rep. 7 (2014), 53-61.

24. Mattarocci S, Shyian M, Lemmens L, Damay P, Altintas DM, Shi T, Bartholomew CR, Thomä NH, Hardy CF, Shore D. Rif1 controls DNA replication timing in yeast through the PP1 phosphatase Glc7. Cell Rep. 7 (2014), 62-69.

25. Yamazaki S, Hayano M, Masai H. Replication timing regulation of eukaryotic replicons: Rif1 as a global regulator of replication timing. Trends Genet. 29 (2013), 449-460.

26. Foti R, Gnan S, Cornacchia D, Dileep V, Bulut-Karslioglu A, Diehl S, Buness A, Klein FA, Huber W, Johnstone E, Loos R, Bertone P, Gilbert DM, Manke T, Jenuwein T, Buonomo SC. Nuclear Architecture Organized by Rif1 Underpins the Replication-Timing Program. Mol Cell. 61 (2016), 260-273.

27. Fernandez-Vidal A, Guitton-Sert L, Cadoret JC, Drac M, Schwob E, Baldacci G, Cazaux C, Hoffmann JS. A role for DNA polymerase θ in the timing of DNA replication. Nat Commun. 5 (2014), 4285.

28. Klein KN, Zhao PA, Lyu X, Sasaki T, Bartlett DA, Singh AM, Tasan I, Zhang M, Watts LP, Hiraga SI, Natsume T, Zhou X, Baslan T, Leung D, Kanemaki MT, Donaldson AD, Zhao H, Dalton S, Corces VG, Gilbert DM. Replication timing maintains the global epigenetic state in human cells. Science. 372 (2021), 371-378.

29. Kogoma T. Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. Microbiol Mol Biol Rev. 61 (1997), 212-38.

30. Drolet M, Brochu J. R-loop-dependent replication and genomic instability in bacteria. DNA Repair (Amst). 84 (2019), 102693.

31. Hawkins M, Malla S, Blythe MJ, Nieduszynski CA, Allers T. Accelerated growth in the absence of DNA replication origins. Nature. 503 (2013), 544-547.

32. Kreuzer KN, Brister JR. Initiation of bacteriophage T4 DNA replication and replication fork dynamics: a review in the Virology Journal series on bacteriophage T4 and its relatives. Virol J. 7 (2010), 358.

33. Patel PK, Arcangioli B, Baker SP, Bensimon A, Rhind N. DNA replication origins fire stochastically in fission yeast. Mol Biol Cell. 17(2006), 308-316.

34. Dellino GI, Cittaro D, Piccioni R, Luzi L, Banfi S, Segalla S, Cesaroni M, Mendoza-Maldonado R, Giacca M, Pelicci PG Genome-wide mapping of human DNA-replication origins: levels of transcription at ORC1 sites regulate origin selection and replication timing. Genome Res 23 (2013), 1–11.

35. Mesner LD, Valsakumar V, Cieslik M, Pickin R, Hamlin JL, Bekiranov S Bubble-seq analysis of the human ge- nome reveals distinct chromatin-mediated mechanisms for regulating early- and late-firing origins. Genome Res 23 (2013), 1774–1788.

36. Petryk N, Kahli M, d'Aubenton-Carafa Y, Jaszczyszyn Y, Shen Y, Silvain M, Thermes C, Chen CL, Hyrien O Replication landscape of the human genome. Nat Commun 7 (2016), 10208.

37. Langley AR, Graf S, Smith JC, Krude T Genome-wide identification and characterisation of human DNA replication origins by initiation site sequencing (ini-seq). Nucleic Acids Res 44 (2016),10230–10247.

38. Macheret M, Halazonetis TD Intragenic origins due to short G1 phases underlie oncogene-induced DNA replication stress. Nature 555 (2018), 112–116.

39. Cayrou C, Ballester B, Peiffer I, Fenouil R, Coulombe P, Andrau JC, van Helden J, Méchali M. The chromatin environment shapes DNA replication origin organization and defines origin classes. Genome Res. 25 (2015), 1873-1885.

40. Comoglio F, Schlumpf T, Schmid V, Rohs R, Beisel C, Paro R. High-resolution profiling of Drosophila replication start sites reveals a DNA shape and chromatin signature of metazoan origins. Cell Rep. 11 (2015), 821-834.

41. Valton AL, Hassan-Zadeh V, Lema I, Boggetto N, Alberti P, Saintomé C, Riou JF, Prioleau MN. G4 motifs affect origin positioning and efficiency in two vertebrate replicators. EMBO J. 33 (2014), 732-746.

42. Masai H, Kanoh Y, Kakusho N, Fukatsu R. Detection of cellular G-quadruplex by using a loop structure as a structural determinant. Biochem Biophys Res Commun. 531 (2020), 75-83.