Cell, Vol. 82, 535–542, August 25, 1995, Copyright © 1995 by Cell Press

Prokaryotic and Eukaryotic Replicons

Meeting Review

Joel A. Huberman
Department of Molecular and Cellular Biology
Roswell Park Cancer Institute
Buffalo, New York 14263

The replicon model, proposed by Jacob and Brenner (1963) to explain the regulation of prokaryotic DNA replication, has proved remarkably robust. Despite the many variations in initiation mechanism discovered since 1963, the replicon model may be applicable, with minor modifications, to regulation of DNA replication in all organisms.

For several years, the Replicon Club of Paris, a group of French scientists interested in regulation of replication, has been meeting to discuss new results. This year, with support from the French Centre National de la Recherche Scientifique (CNRS), the Replicon Club sponsored an international Jacques Monod Conference on Prokaryotic and Eukaryotic Replicons. The meeting was organized by G. Buttin (Institut Pasteur, Paris) with assistance from A. Falaschi (International Centre for Genetic Engineering and Biotechnology, Trieste, Italy) and M. Kohiyama (Institut Jacques Monod, Paris) and was held at the CNRS's Centre Paul Langevin in the Alpine village of Aussois, France, from June 18–22, 1995. There, invigorated by beautiful surroundings and excellent cuisine, scientists from around the world provided evidence both for the vitality of the replicon model and for substantial advances in our understanding of the regulation of replication.

Variety of Replicons

The original replicon model (Figure 1) suggested that replication is positively controlled by an initiator protein that acts on a single replicator to initiate replication of a circular replicon. Extension of the model to other organisms requires allowance for multiple linear chromosomes in eukaryotic cells, each possibly having multiple replicators (Figure 1). This extension requires that replicon be redefined to mean the stretch of DNA replicated from a single replicator. An updated model also requires allowance for the possible existence of multiple initiator proteins, which may form a complex (as in Figure 1) or may separately bind to different portions of the replicator, and recognition that interactions between initiators and replicators are typical binding reactions, governed by the laws of mass action (Figure 1). Consequently, if the concentration of initiators is high or their specificity is low, most or all DNA sequences may be able to serve as replicators.

At the meeting in Aussois, developments relating to a wide variety of replicons were presented.

Protein Priming

The Bacillus subtilis phage, Φ29, has a linear double-stranded DNA with a viral terminal protein (TP) covalently attached to its 5′ ends. L. Blanco (Centro de Biología Molecular Severo Ochoa, Madrid) described the mechanism of initiation of Φ29 replication. For each round, new TPs, complexed with viral DNA polymerase molecules, are positioned at the 3′ ends of the parental duplex. Using the penultimate 3′ deoxyribosylthymine (dT) as template, the polymerase then catalyzes the formation of a covalent bond between deoxyribosyladenine (dA) and a serine of TP. Next, the TP-dA slides back 1 nt to position the dA adjacent to the 3′ terminal dT of each strand. The polymerase then catalyzes processive elongation of the new protein-primed strands, using the 3′-ended strands as templates and displacing the 5′-ended strands. This "sliding-back" initiation mechanism, which provides an opportunity during subsequent replication rounds to correct nucleotides incorrectly incorporated at the first step, appears to be a common feature of linear replicons employing the TP mechanism.

Rolling Circle

S. D. Ehrlich (Institut National de la Recherche Agronomique [INRA], Jouy en Josas, France) discussed the Rep protein of pC194, a rolling circle plasmid of gram-positive bacteria. A major difference between pC194 and the well-studied rolling circle replicon ΦX174 is that ΦX174 is a virus with runaway replication, while pC194 is a plasmid with regulated replication. The difference can be explained by comparison of the two Rep proteins. Both proteins initiate replication by making nicks at specific sites in one of the DNA strands, yielding 3′ ends that serve as primers and 5′ ends covalently attached to a Rep protein tyrosine. After one round of synthesis, second nicks are made at the same sites. The ΦX174 Rep protein employs a second tyrosine to catalyze the second nick, permitting retention of the 5′ end and, ultimately, reinitiation. In contrast, pC194 appears to use glutamic acid–catalyzed hydrolysis for the second nick, thus losing bond energy and preventing reinitiation.

Rolling circle replication is also employed by geminiviruses, whose small circular single-stranded genomes replicate via double-stranded intermediates in the nuclei of plant cells. Since replication is catalyzed entirely by enzymes from the host cell (except the viral Rep protein) and viral DNA is packed into chromatin, study of geminivirus replication appears to be a promising way to learn more about DNA replication and the cell cycle in plants. B. Gronenborn (Institut des Sciences Végétales, Gif sur Yvette, France) reported that geminivirus Rep proteins nick the origin via a tyrosine and have a DNA-independent ATPase/GTPase activity that is essential for in vivo replication and reminiscent of the GTPase in signal transduction G proteins. C. Gutierrez (Centro de Biología Molecular Severo Ochoa, Madrid) showed that geminivirus Rep contains an LXCXE motif capable of mediating stable binding to the human retinoblastoma (Rb) protein. This motif is essential for in vivo replication, suggesting that viral replication is linked to cell cycle control and that plants, like animals, may contain Rb homologs that inhibit entry into S phase.

Unidirectional θ

C. Bruand and E. Le Chatelier (INRA, Jouy en Josas, France) described the pAMβ1 replicon, a plasmid of gram-

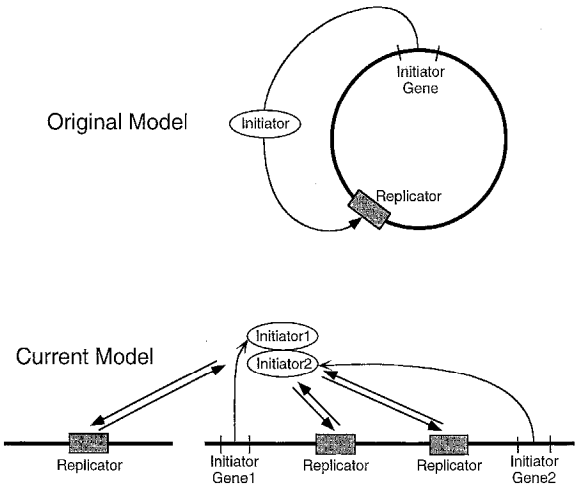

Figure 1. Past and Present Replicon Models

(Top) Model proposed by Jacob and Brenner (1963). (Bottom) Model arising from discussions at the Aussois meeting described in this review.

positive bacteria, which bears some resemblance to the well-known ColE1 replicon. For pAMβ1, a plasmid-encoded positive regulator, the *RepE* gene product, is essential for replication, and DNA polymerase I is replaced by DNA polymerase III holoenzyme (HpolIII) about 200 bp downstream of *ori*. According to M.-A. Petit and L. Janniere (INRA, Jouy en Josas, France), this polymerase switch is aided by the pAMβ1 resolvase, which binds tightly about 250 bp downstream of *ori*, thereby blocking polymerase I and creating a D loop with an exposed primosome assembly site where an HpolIII-based replication fork can efficiently be set up. V. Bidnenko (INRA, Jouy en Josas, France) showed that the plasmid-encoded topoisomerase, topβ, also assists by relaxing plasmid DNA and thereby removing the driving force for strand unwinding when polymerase I reaches about 190 bp downstream of *ori*. The presenters proposed that similar mechanisms may facilitate polymerase switching in ColE1 and eukaryotic replicons.

Previously studied prokaryotic and eukaryotic circular plasmids containing two replication origins utilize only one origin at a time. J. B. Schvartzman (Centro de Investigaciones Biológicas, Madrid) reported that occasionally molecules of the plasmid pPI21, containing two unidirectional ColE1 origins in opposite orientation, employ both origins at once, resulting in stable "bubble" structures detectable by two-dimensional gel electrophoresis and electron microscopy.

Bidirectional θ Early, Rolling Circle Later

R. Skaliter (Stanford University, California) described similarities between herpes virus replication and λ replication. In both cases, early origin-dependent θ-mode replication is followed by subsequent rolling circle replication to generate concatemers that are packaged into viral particles. In the case of herpesvirus, two-dimensional gel electrophoretic evidence suggests origin-specific initiation mediated by the viral origin–binding protein UL9 as early as 2 hr after infection. Evidence for interactions between UL9 and cellular DNA polymerase α, but not the viral DNA polymerase, suggests that initiation at viral origins may be accomplished by UL9 and cellular polymerases. Later rolling circle replication is independent of UL9 but requires the other viral replication proteins, which, according to physical and electron microscopic evidence, may exist as a complex.

Gel shift and footprinting experiments described by P. Elias (Göteborg University, Sweden) suggest that a protein complex may also function at the viral origin. The complex would consist of two dimers of UL9, interacting through their C-terminal domains both with specific binding sites in the origin and with four viral single-stranded DNA-binding proteins, which are called ICP-8. Since ICP-8 binds more tightly to single-stranded DNA than to UL9, it is likely that once single-stranded DNA is exposed at the origin by unwinding of parental strands, ICP-8 is transferred to the single-stranded DNA.

During the early stage of bacteriophage λ replication, initiation is mediated by the phage-encoded O protein at *oriλ*. Under in vivo conditions, transcription near *oriλ*, usually from the P_R promoter, is also required to initiate replication. G. Wegrzyn (University of Gdansk, Poland) discussed circular molecules derived from λ but missing most of the λ genome except the origin region. He presented evidence that O protein, along with other proteins involved in initiation, is retained at one of the two daughter origins during each replication round. Such inherited replication complexes may mediate further replication rounds. Runaway replication is prevented by the requirement that each round be activated by transcription at P_R. Such transcription is mediated by the host dnaA protein, which is present in high concentration during only a limited portion of each cell cycle.

Bidirectional θ

The single initiator protein of Escherichia coli, encoded by the *dnaA* gene, governs initiation at the single replicator, *oriC* (reviewed by Kornberg and Baker, 1991). Interactions between the dnaA protein and *oriC* were described by W. Messer (Max-Planck-Institut für Molekulare Genetik, Berlin). Within *oriC* there are four 9 nt repeats, called dnaA boxes, that serve as specific binding sites for the dnaA protein. Measurements of binding to the individual boxes and to complete *oriC* revealed that binding is highly context dependent. The dnaA protein does not contain any of the known DNA-binding motifs. Protein fusions and mutational analysis revealed that the C-terminal 94 amino acids, including three α helices, are responsible for DNA binding.

Initiation of SV40 DNA replication requires binding of two back-to-back hexameric complexes of the SV40 initiator/helicase T antigen to the SV40 replication origin. E. Fanning (Vanderbilt University, Nashville, Tennessee) described the importance of phosphorylation at Thr-124 of T antigen for cooperative interactions between the two hexamers, which are essential for processive bidirectional DNA unwinding from the origin. Similar interactions between helicases may be important for unwinding at cellular origins.

In vivo, both of the bovine papilloma virus (BPV) E1 and E2 proteins are essential for initiation of replication. Like SV40 T antigen and herpes UL9, E1 is an origin-binding protein with helicase activity. E2 is a transcription factor capable of forming a complex with E1. P. Clertant (University of Nice, France) reported that his laboratory has developed an origin- and E1-dependent, but E2-independent, BPV in vitro replication system as efficient as that for SV40. In vitro E2 appears to only inhibit nonspecific initiation, but in vivo it may assist in "opening" chromatin at the origin. Since the BPV in vitro system catalyzes multiple replication rounds, additional factors, possibly of cellular origin, may be responsible for limiting BPV replication in vivo to an average of one round per cell cycle.

Epstein–Barr virus (EBV) can establish a latent infection of human B lymphocytes in which each large circular viral DNA molecule is replicated once per cell cycle. Previous investigations employing plasmids containing small fragments of the EBV genome had suggested that EBV latent replication depends on an origin sequence, *oriP*, a virally encoded protein (EBNA1), and cellular replication proteins (reviewed by Yates, 1993). R. Little (Albert Einstein College of Medicine, New York) reported that replication of large EBV genomes initiates both at *oriP* and also in broader zones located elsewhere in the EBV genome. This combination of specific initiation and initiation in broader zones is similar to initiation in mammalian chromosomes (see below).

Bidirectional Linear

Replication of Saccharomyces cerevisiae chromosomes is largely consistent with the current replicon model (Figure 1). Binding of an initiator protein complex (ORC) to a replicator (autonomously replicating sequence [ARS] element) is essential for initiation at or near the ARS element (reviewed by Newlon and Theis, 1993). ARS elements are identified by their ability to serve as replication origins in plasmids. Interestingly, some ARS elements do not serve as origins in the chromosome. M. Weinberger (Roswell Park Cancer Institute, Buffalo, New York) reported that one such chromosomally inactive ARS element, *ARS301*, contains the sequence motifs found in typical chromosomally active ARS elements: an 11 bp ARS consensus sequence and an essential flanking sequence. J. Diffley (Imperial Cancer Research Fund, Clare Hall Laboratories, South Mimms, England) added that *ARS301*, in its normal location on chromosome III, yields prereplicative and postreplicative footprints (see below) that are similar to those of chromosomally active ARS elements. Thus, the reasons for the chromosomal inactivity of *ARS301* remain obscure.

In the yeasts Schizosaccharomyces pombe and Yarrowia lipolytica, chromosomal DNA replication also is dependent on and initiates at or near defined sequence elements, as described, respectively, by J. Huberman (Roswell Park Cancer Institute, Buffalo, New York) and P. Fournier and L. Vernis (INRA, Thiverval-Grignon, France). In S. pombe, the number of sequence motifs contributing to ARS function (two essential motifs and more than eight stimulatory motifs in the case of *ars3002*) is larger than in S. cerevisiae. In Y. lipolytica, the origins contain no obvious consensus sequence, are considerably more GC rich than S. cerevisiae and S. pombe ARS elements, and do not support extrachromosomal replication of plasmids lacking centromeres.

The slime mold Physarum polycephalum can exist as a giant single cell containing millions of naturally synchronous nuclei. G. Pierron (CNRS, Villejuif, France) took advantage of this synchrony to demonstrate that replication initiates in or near the promoters of four genes that replicate in early S phase, suggesting the presence of specific replicators. M. Bénard's (CNRS, Villejuif, France) demonstration that allelic origins are activated simultaneously also suggests specific replicators.

To What Extent Does the Replicon Model Apply to Animal Cell DNA Replication?

For the past 5 years, investigators of DNA replication in animal cells have been facing a paradox. Some results indicate that replication initiates at specific locations, and other observations suggest that replication initiates randomly within broad initiation zones (reviewed by Coverley and Laskey, 1994). Similarly conflicting results were presented in Aussois. A. Falaschi (International Centre for Genetic Engineering and Biotechnology, Trieste, Italy) described the presence of short nascent strands, indicative of a replication origin, in all tested human cell lines within an ~500 bp region at the 3′ end of the human lamin B2 gene. K.-I. Tsutsumi (Iwate University, Ueda, Japan) observed that an ~900 bp fragment encompassing the promoter of the rat aldolase B gene is capable of promoting autonomous replication in cultured cells. M. Zannis-Hadjopoulos (McGill University, Montreal) reviewed studies from her lab that indicate, by several methods, that many of the short nascent strands generated in earliest S phase are associated with chromosomal replication origins and are themselves capable of stimulating replication in vivo and in vitro (see below). G. Wahl (Salk Institute, La Jolla, California) provided evidence, obtained with several techniques, for initiation within discrete loci near the human β-globin gene (see below) and, interestingly, within the Syrian hamster *CAD* gene. The *CAD* gene is transcribed in early S phase, overlapping the time at which it is replicated.

Several other investigators described evidence for broad initiation zones. P. Dijkwel (University of Virginia, Charlottesville) detected a broad initiation zone encompassing the transcriptionally silent rhodopsin gene in Chinese hamster ovary (CHO) cells. J. Hamlin (University of Virginia, Charlottesville) found that initiation downstream of the dihydrofolate reductase (DHFR) gene in CHO cells is distributed over a broad zone regardless of whether the region is amplified. Despite the presence of multiple potential initiation sites, the zone as a whole is not efficient but is frequently replicated passively by forks entering from flanking regions. R. Little (Albert Einstein College of Medicine, New York) reviewed his observation that replication initiates at multiple sites within the nontranscribed spacer of human rDNA, and M. Debatisse-Buttin (Institut Pasteur, Paris) described evidence for initiation within broad nontranscribed stretches in a multigenic region near the mammalian *AMPD2* gene. M. Calos (Stanford Univer-

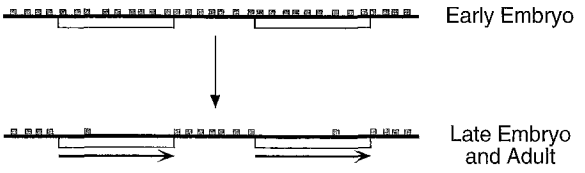

Figure 2. The Transition from Random to Nonrandom Initiation of Replication Coincides with the Beginning of Embryonic Transcription during Early Embryogenesis

The diagram summarizes results presented by O. Hyrien (Institut Jacques Monod, Paris) concerning initiation of replication in Xenopus rDNA, but the general concept may be applicable to the rest of the genome. The small stippled boxes represent initiation sites, the large open boxes are genes, and the horizontal arrows represent transcripts.

sity, California) provided evidence for sequence-independent plasmid replication in human and Drosophila cells.

Attempts to rationalize these apparently contradictory observations were the subject of extensive discussions. All participants agreed that the evidence for sequence-independent replication during early Xenopus embryogenesis is convincing (reviewed by Coverley and Laskey, 1994) and that initiation is not completely random in adult cells (it occurs at specific sites or in broad initiation zones, but not everywhere). O. Hyrien (Institut Jacques Monod, Paris) presented data indicating that the developmental transition from completely random initiation in Xenopus rDNA to initiation primarily within the nontranscribed spacer starts at the midblastula stage, the time when rDNA transcription begins (Figure 2). This observation suggests that the chromatin restructuring necessary to limit transcription to certain regions may also serve to limit initiation of replication to certain regions.

How can these observations be accommodated by the replicon model? Two nonmututally exclusive hypotheses were proposed. M. Méchali (Institut Jacques Monod, Paris) suggested that specificity is lost in the early embryo because of the high concentration of maternally derived initiation proteins. Reduction of initiator concentration and establishment of transcriptionally active chromatin (Figure 2) at the midblastula transition may explain the appearance of initiation zones. M. Calos (Stanford University, California) proposed that there are no rigorous sequence requirements for initiation even in adult cells. If an initiator is required, that initiator must have low sequence specificity. Potential initiation sequences would have to be present at a frequency of one or more per kilobase, but most of them would be repressed by chromatin structure. Transcription is a major factor in determining chromatin structure and therefore would play a key role in specifying which sites would be used.

There was general agreement that further progress in understanding animal cell replication origins requires genetic experiments to identify precisely the *cis*-acting sequences responsible for initiation sites and initiation zones. Once these sequences are identified, it should be possible to identify interacting proteins and determine whether those proteins serve as true initiators, serve to establish transcriptional patterns, or function in some other way.

Genetic Experiments Suggest Multiple Determinants of Mammalian Replication Origins

The results of two recent genetic experiments on mammalian origins were described at the meeting. These and previously reported experiments are summarized in Figure 3.

An initiation zone of more than 55 kb is located downstream of the DHFR gene in CHO cells. Within this zone, regions called oriβ and oriγ initiate replication at relatively high frequency (reviewed by Hamlin et al., 1994). J. Hamlin (University of Virginia, Charlottesville) has found that a 75 kb deletion upstream of the DHFR gene that includes the DHFR promoter and thereby prevents DHFR transcription leads to loss of detectable initiation within the zone. Previously, Handeli et al. (1989) reported that when a 16 kb stretch of DNA containing oriβ (λD6A; Figure 3A) is transplaced to other locations in the CHO genome, oriβ, sequences near it, or both continue to initiate replication. Considering the new data from J. Hamlin, it is difficult to understand the Handeli et al. (1989) result unless transplacement was favored to regions that were also favorable for replication initiation. Repetition of that experiment with improved characterization of the locations of the transplaced λD6A segments would be useful.

Both Kitsberg et al. (1993) and G. Wahl (Salk Institute, La Jolla, California) have obtained evidence for a narrow initiation zone or specific initiation site near the 5′ end of the human β-globin gene (Figure 3B, normal). Kitsberg et al. (1993) found that a deletion covering the preferred initiation site eliminated initiation in the region, leading to

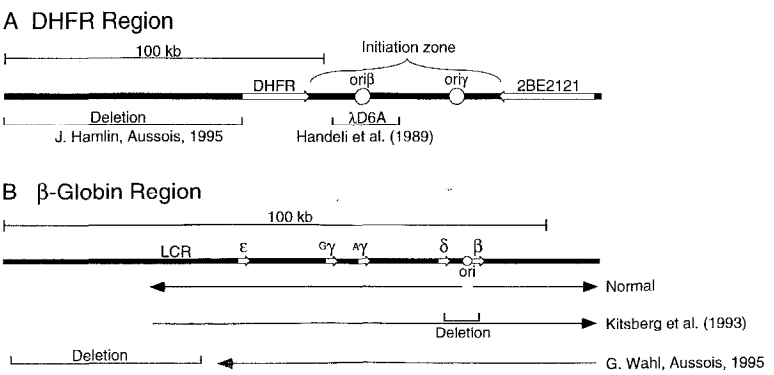

Figure 3. Effects of Deletions and a Transplacement on Mammalian Origin Function

The open arrows represent genes, and the open circles represent sites or regions of preferred initiation. The long, thin arrows in (B) represent preferred directions of replication fork movement. LCR, locus control region.