Cell, Vol. 52, 635–638, March 11, 1988, Copyright © 1988 by Cell Press

Transcriptional Elements As Components of Eukaryotic Origins of DNA Replication

Minireview

Melvin L. DePamphilis
Department of Cell and Developmental Biology
Roche Institute of Molecular Biology
Nutley, New Jersey 07110

Genetically defined *cis*-acting sequences that function as origins of DNA replication (*ori* regions) in eukaryotic genomes frequently contain two primary components: a core component, dedicated to DNA replication, that is required for replication under all conditions, and an auxiliary component, containing promoter or enhancer elements, that may be involved in transcription as well as replication and is dispensable under some conditions. The core component determines where replication begins, and the auxiliary component determines efficiency of replication and/or cell specificity. Interactions between core and auxiliary components can be divided into three groups (see figure). In Group I, an enhancer or promoter can reside in either orientation relative to *ori*-core as long as it lies close to the AT-rich end, and in this way facilitates interaction between *ori*-core and proteins that unwind DNA templates and initiate DNA synthesis. In Group II, enhancer and *ori*-core function independently of their relative orientation and distance from one another, as though *ori*-core subverts the enhancer into activating initiation of DNA replication rather than transcription. In Group III, mRNA passing through *ori*-core from an upstream promoter is converted into RNA primers for DNA synthesis. The varied ways in which *cis*-acting transcriptional elements facilitate initiation of replication in these three groups are discussed below.

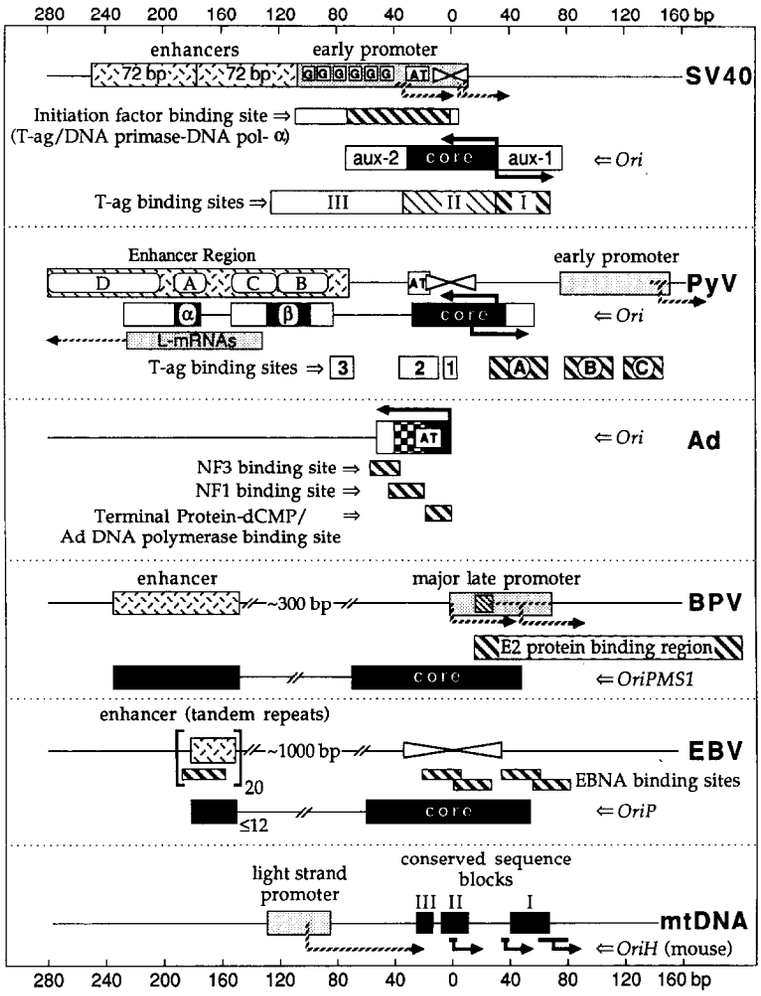
Group I—SV40, Polyoma Virus, and Adenovirus

SV40 and polyoma virus (PyV) *ori*-core components share common sequence motifs, and initiate replication only in the presence of large tumor antigen (T-ag) encoded by the same virus and DNA primase–DNA polymerase α from permissive cells. T-ag has helicase activity that unwinds double-stranded DNA at *ori*-core (Stahl et al., EMBO J. *5*, 1939–1944, 1986; Dean et al., PNAS *84*, 8267–8271, 1987; Dodson et al., Science *238*, 964–967, 1987). RNA-primed DNA synthesis is then initiated exclusively on the template strand encoding early mRNA at one of several possible initiation sites (Hendrickson et al., EMBO J. *6*, 2011–2018, 1987; Taljanidisz et al., NAR *15*, 7877–7888, 1987; Decker et al., MCB *6*, 3815–3825, 1986). Transition sites on each strand of *ori* where RNA-primed initiation events stop and continuous DNA synthesis begins (the origin of bidirectional DNA replication) occur in *ori*-core (solid arrows in figure) at virtually the same positions in SV40 and PyV relative to sequence motifs that are required for T-ag binding and DNA replication. Subsequent initiation of RNA-primed nascent DNA chains (Okazaki fragments) occurs only on the retrograde arm of replication forks (Hendrickson et al., NAR *15*, 6369–6385, 1987). Thus, the *ori*-cores from SV40 and PyV appear to initiate replication by the same mechanism. However, PyV requires an enhancer element upstream of the first initiation event in *ori*-core, whereas SV40 *ori*-core does not.

An enhancer element is required for PyV *ori*-dependent replication in developing two-cell mouse embryos, in mouse embryonic and differentiated cell lines, and in the living animal (Wirak et al., MCB *5*, 2924–2935, 1985; Amati, Cell *43*, 561–562, 1985; Veldman et al., MCB *5*, 649–658, 1985; Hassell et al., In Cancer Cells, Vol. 4, eds. Botchan et al., Cold Spring Harbor Laboratory, 1986, pp. 561–569; Campbell and Villarreal, MCB *6*, 2068–2079, 1986; Rochford et al., PNAS *84*, 449–453, 1987; Tang et al., MCB *7*, 1681–1690, 1987). Enhancer function is normally provided by either the α or β element (the solid boxes are the minimal sequences), corresponding to enhancer domains A and B + C, respectively. In addition, non-PyV enhancers can substitute for α and β, and *cis*-acting mutations in these sequences can allow PyV *ori*-core to initiate replication in mouse cell types that are normally nonpermissive. As it does in activation of promoters, the enhancer element determines cell-type specificity for activation of PyV *ori*-core, but it does not alter PyV *ori*-core specificity for mouse cells (Campbell and Villarreal, MCB *5*, 1534–1537, 1985). The ability of enhancer elements to stimulate replication from *ori*-core correlates closely with their ability to stimulate transcription from promoters, although the PyV *ori*-core is about 3-fold more responsive than the β-globin promoter.

SV40 *ori*-core also responds to transcriptional elements, but to a lesser extent than PyV *ori*-core. Deletion of the 72 bp repeats (enhancers) has no effect on DNA replication as long as three or more of the six $G_{3-4}CG_2$ motifs (G boxes) that make up part of the promoter are present (the *aux-2* element). Elimination of all G boxes reduces DNA replication in vivo 2- to 3-fold, but elimination of all G boxes and enhancer elements reduces replication 10- to 100-fold, depending on the presence or absence of *aux-1* and other experimental conditions (DeLucia et al., J. Virol. *57*, 138–144, 1986; Li et al., MCB *6*, 1117–1128, 1986; Hertz and Mertz, MCB *6*, 3513–3522, 1986). Enhancer and promoter elements function in either orientation, but they must be in close proximity to the AT-rich side of SV40 or PyV *ori*-core (Innis and Scott, MCB *4*, 1499–1507, 1984; Lee-Chen and Woodworth-Gutai, MCB *6*, 3086–3093, 1986; Chandrasekharappa and Subramanian, J. Virol. *61*, 2973–2980, 1987; Hassell et al., op. cit.). Therefore, the PyV and SV40 *ori*-core components are affected by transcriptional elements in a manner that is qualitatively, but not quantitatively, the same.

At first glance, initiation of adenovirus (Ad) DNA replication appears quite different from the corresponding events in SV40 and PyV. Replication begins at each end of the linear Ad genome by the binding of an Ad terminal protein–dCMP complex, which functions as a primer for the initiation of DNA synthesis by Ad DNA polymerase. One strand is then replicated completely while the other strand is displaced; Okazaki fragments are not synthesized on the retrograde arm of the fork. Nevertheless, like the SV40



and PyV *ori* regions, the Ad *ori* contains both promoter and enhancer sequence motifs. The terminal 18 bp core sequence, required for DNA replication in all adenovirus strains, is the binding site for an Ad terminal protein–dCMP/Ad DNA polymerase initiation complex. In one viral strain (Ad2), interaction of this initiation complex with *ori*-core is facilitated 10-fold by a cellular protein called nuclear factor I (NF-I) and another 3-fold by NF-III binding to the adjacent 34 nucleotides (Wides et al., MCB *7*, 864–874, 1987; Rosenfeld et al., MCB 7, 875–886, 1987; Pruijn et al., EMBO J. *6*, 3771–3778, 1987). Both *ori*-core (solid box) and the NF-I binding site (checkered box) are required for replication in vivo (Hay EMBO J. *4*, 421–426, 1985). NF-I is indistinguishable from transcriptional factor CTF, which binds to the CCAAT motif and stimulates transcription from several promoters (Jones et al., Cell *48*, 79–89, 1987), and the NF-III binding site is found in several promoter and enhancer regions. As is the case for SV40 and PyV, the distance but not the orientation of the transcriptional element (NF-I binding site) relative to the AT-rich end of *ori*-core is critical (Adhya et al., JBC *261*, 3339–3346, 1986; Wides et al., op. cit.).

The requirement for auxiliary components in the Ad *ori* depends on the source of DNA replication proteins. Ad4 *ori*-core replicates almost as well in vivo as the complete Ad4 *ori* (Hay, JMB *186*, 129–136, 1985). Ad4 *ori* is just as efficient as Ad2 *ori* even though AD4 *ori* lacks the NF-I binding site while retaining the NF-III site. Not surprisingly, Ad4 replication in vitro does not require NF-I, although it does respond to NF-III. Thus, Ad *ori*-core alone is sufficient for replication under some conditions. The difference between the Ad2 and Ad4 *ori* regions reflects a difference in affinity of the Ad terminal protein–dCMP/DNA polymerase initiation complex for the *ori*-core. Ad4 replication proteins work equally well with either Ad2 or Ad4 *ori*, whereas Ad2 replication proteins work well only with Ad2 *ori*, revealing a strong dependence for their interaction with *ori*-core on concomitant binding of NF-I. In this sense the Ad2 *ori* is analogous to the PyV *ori*: interaction of DNA replication proteins with *ori*-core is strongly dependent on the presence of a transcriptional element. Conversely, the Ad4 *ori* is analogous to the SV40 *ori*.

How do transcriptional elements facilitate *ori*-core function? As in Ad, *ori*-cores from PyV and SV40 contain all the

cis-acting information necessary to initiate replication since they can function as well as complete *ori* regions in some in vitro DNA replication systems and following their injection into pronuclei of fertilized mouse eggs (Li et al., op. cit.; Prives et al., MCB *7*, 3694–3704, 1987; DePamphilis et al., In Cancer Cells, Vol. 6, eds. Stillman and Kelly, Cold Spring Harbor Laboratory, in press). Therefore, as demonstrated with the Ad *ori*, transcriptional elements must facilitate binding of replication proteins to *ori*-core. This is not accomplished by promoting transcription through *ori*-core. Initiation of SV40 and Ad DNA replication is resistant to α-amanitin (Decker et al., JBC *262*, 10863–10872, 1987; Li and Kelly, PNAS *81*, 6973–6977, 1984), all PyV mRNAs (cross-hatched arrows) are initiated outside *ori*-core and extend away from *ori*, and Ad *ori* is not involved in Ad transcription.

One hypothesis for the role of transcriptional elements in *ori* function is that some proteins that bind specifically to transcriptional elements may interact directly with proteins binding to *ori*. For example, the AP-2 protein binds to T-ag as well as to the SV40 enhancer and *aux-2* sequences (Mitchell et al., Cell *50*, 847–861, 1987). The *aux-2* element is part of the normal binding site for SV40 initiation factors (Yamaguchi and DePamphilis, PNAS *83*, 1646–1650, 1986), and *ori*$^-$ mutations in the AT-rich motif of *ori*-core can be suppressed by alterations in *aux-2* (Gerard and Gluzman, MCB *6*, 4570–4577, 1986). Deletion of all six G boxes shifts the beginning of continuous DNA synthesis on the early mRNA side of *ori* about 40 bp downstream, consistent with a change in the initiation factor binding site (DePamphilis et al., op. cit.). Alternatively, proteins recognizing transcriptional elements may stabilize binding of initiation proteins to *ori*-core by promoting localized strand separation. Either mechanism would account for the required proximity between the auxiliary and core components of *ori* in these viruses.

A second hypothesis is that proteins that bind to promoters and enhancers modify the chromatin structure of *ori*-core, making it more accessible to the T-ag initiation complex. For example, TATA box binding protein TFIID can prevent nucleosome-mediated repression of a promoter (Workman and Roeder, Cell *51*, 613–622, 1987). Similarly, G boxes and, to a lesser extent, 72 bp repeats are responsible for creating nuclease-hypersensitive sites in SV40 *ori*-core (Gerard et al., MCB *5*, 52–58, 1985; Jongstra et al., Nature *307*, 708–714, 1984). Thus, conditions under which transcriptional elements are not needed to facilitate replication could represent high concentrations of initiation proteins, altered initiation proteins, and/or absence of normal chromatin structure. PyV *ori*-core may be more strongly dependent on *cis*-acting transcriptional elements than SV40 *ori*-core, because T-ag binds weakly to PyV *ori*-core (sites 1 and 2) but strongly to SV40 *ori*-core (site II); the major PyV T-ag binding sites A, B, and C lie outside of *ori* (see figure).

Group II—Bovine Papilloma Virus and Epstein-Barr Virus

Auxiliary components and *ori*-core can also be independent of distance as well as orientation. Bovine papilloma virus (BPV) plasmid maintenance sequence 1 (*oriPMS1*) consists of two *cis*-acting components (Lusky and Botchan, PNAS *83*, 3609–3613, 1986). One component appears to be an enhancer because it functions independently of orientation and position with respect to the second sequence and can be replaced by known enhancers such as the Harvey sarcoma virus long terminal repeat and the PyV β element. The second component overlaps a promoter (Stenlund et al., Science *236*, 1666–1671, 1987; Baker and Howley, EMBO J. *6*, 1027–1035, 1987) and a *cis*-acting region that responds to a protein(s) encoded by open reading frame E2 (Spalholz et al., J. Virol. *61*, 2128–2137, 1987). Open reading frames E2 and E1 are essential for extrachromosomal maintenance. One product of E2 *trans*-activates certain BPV promoters and binds to $ACCN_6GGT$, a motif present in *ori* component 2 (hatched box within promoter). Whether this binding activates *ori* directly or indirectly by activating downstream genes (e.g., E1) is not known. Thus, by analogy with SV40 and PyV, component 2 represents *ori*-core. The presence of a promoter within BPV *ori*-core may allow interaction with the enhancer.

Interaction between an enhancer and *ori*-core is evident in Epstein-Barr virus (EBV). *oriP* contains two *cis*-acting components separated by about 1000 bp, both of which are required for replication of recombinant plasmids in cells that express EBNA, a viral encoded nuclear antigen required for EBV DNA replication (Reisman et al., MCB *5*, 1822–1832, 1985). Neither a precise distance nor a particular orientation of the two components relative to one another is required for *oriP* activity. One component consists of a family of elements repeated in tandem that can function as a transcriptional enhancer when activated by EBNA; twelve copies are sufficient for replication (Reisman and Sugden, MCB *6*, 3838–3846, 1986). EBNA binds to each element, and mutations in EBNA that effect its ability to activate *oriP* also affect its ability to activate a promoter (Rawlins et al., Cell *42*, 859–868, 1985; Yates and Camiolo, In Cancer Cells, Vol. 6, eds. Stillman and Kelly, Cold Spring Harbor Laboratory, in press). Since the SV40 enhancer does not substitute for the EBV enhancer, EBNA binding to both enhancer and *ori*-core appears necessary to initiate replication. The second component, containing a 64 bp palindrome (indicated in the figure by a "bow tie") and three EBNA binding sites, is reminiscent of the SV40, PyV, and herpes simplex virus *ori*-cores, although this component, like component 2 of *oriPMS1*, has not yet been identified as the place where DNA replication begins. It should be noted that the two herpes simplex virus *ori* regions that appear similar to the EBV "core" do not appear to involve *cis*-acting transcriptional elements (Stow et al., In Cancer Cells, Vol. 4, eds. Botchan et al., Cold Spring Harbor Laboratory, 1986, pp. 497–507; Polvino-Bodnar et al., J. Virol. *61*, 3528–3535, 1987).

Group III—Mitochondrial DNA

Mitochondrial DNA (mtDNA) defines a third *ori* group, in which transcription from an upstream promoter appears to provide RNA primers for initiation of DNA synthesis. Initiation of heavy strand mtDNA synthesis of *oriH* occurs by formation of a displacement loop (D loop). RNA-primed DNA synthesis begins at three major sites on the light

strand template, which correlate with three highly conserved DNA sequence blocks, while the heavy strand template is displaced (Chang and Clayton, PNAS *82*, 351–355, 1985; Chang et al., EMBO J. *4*, 1559–1567, 1985). RNA transcripts originating from the light strand promoter begin at a single nucleotide, but some terminate within the D loop at sites corresponding closely to the locations of some 5′ ends of nascent heavy strand DNA. Moreover, the 5′ ends of some heavy strand DNA chains are found covalently attached to an RNA whose 5′ end maps to the light strand promoter's initiation site. Mitochondria also contain an endoribonuclease that cleaves single-stranded RNA at the site where RNA–DNA junctions are observed between conserved sequence blocks II and III (Chang and Clayton, EMBO J. *6*, 409–417, 1987). This enzyme may be involved either in creating or removing RNA primers. Development of in vitro systems that initiate mtDNA synthesis should clarify the mechanism for *oriH* activation.

Cellular Origins of Replication

The virus and plasmid *ori* regions described above suggest an evolutionary relationship between origins of transcription and origins of replication, and portend a role for *cis*-acting transcriptional elements in the initiation of cellular DNA replication. Yeast autonomous replication sequences (ARSs) are examples of cellular chromosomal sequences that can function as origins of DNA replication in yeast plasmids and are thought to serve the same function in chromosomes (Brewer and Fangman, Cell *51*, 463–471, 1987; Huberman et al., Cell *51*, 473–481, 1987). Unfortunately, attempts to isolate cellular *ori* regions by cloning ARS activities may uncover only those in Groups I and III, because the core and auxiliary components of Group II *ori* regions may be separated accidently. Nevertheless, the "silencer" sequence (the opposite of an enhancer) involved in repression of yeast mating type loci includes an ARS element as one of its three components, suggesting a functional interaction between replication and transcriptional elements in cellular chromosomes (Brand et al., Cell *51*, 709–719, 1987). Such an interaction might explain why transcriptionally active genes are replicated early during S phase of mammalian cells whereas quiescent genes are replicated late (Iqbal et al., NAR *15*, 87–103, 1987), and provide a role for *cis*-acting transcriptional elements in regulation of cellular DNA replication. In addition to *cis*-acting DNA sequences, cellular chromosomes may require organization into a nuclear structure before they can initiate replication (Newport, Cell *48*, 205–217, 1987; Blow and Watson, EMBO J. *6*, 1997–2002, 1987). In contrast, the three lytic virus *ori* regions in Group I can initiate replication in soluble systems with purified proteins, suggesting that they are designed to escape controls that restrict cellular origins to one initiation event per S phase. Thus, Group I origins of replication may represent a mechanism for gene amplification, while Group II and III origins may represent mechanisms for cellular and plasmid replication.