Journal of Cellular Biochemistry 60:297–316 (1996)

Common Structural Features of Replication Origins in All Life Forms

Teni Boulikas

Institute of Molecular Medical Sciences, Palo Alto, California 94306

Abstract Origins of replication (ORIs) among prokaryotes, viruses, and multicellular organisms appear to possess simple tri-, tetra-, or higher dispersed repetitions of nucleotides, AT tracts, inverted repeats, one to four binding sites of an initiator protein, intrinsically curved DNA, DNase I–hypersensitive sites, a distinct pattern of DNA methylation, and binding sites for transcription factors. Eukaryotic ORIs are sequestered on the nuclear matrix; this attachment is supposed to facilitate execution of their activation/deactivation programs during development. Furthermore, ORIs fall into various classes with respect to their sequence complexity: those enriched in AT tracts, those with GA- and CT-rich tracts, a smaller class of GC-rich ORIs, and a major class composed of mixed motifs yet containing distinct AT and polypurine or GC stretches. Multimers of an initiator protein in prokaryotes and viruses that might have evolved into a multiprotein replication initiation complex in multicellular organisms bind to the core ORI, causing a structural distortion to the DNA which is transferred to the AT tract flanking the initiator protein site; single-stranded DNA-binding proteins then interact with the melted AT tract as well as with the DNA polymerase α-primase complex in animal viruses and mammalian cells, causing initiation in DNA replication. ORIs in mammalian cells seem to colocalize with matrix-attached regions and are proposed to become DNase I–hypersensitive during their activation.

Key words: Replication origins, DNA repeats, initiator protein, DNA distortion, inverted repeat, enhancer, silencer, transcription factor, SV40 ORI, HSV-1 ORI nuclear matrix

Unlike *E. coli*, which uses a single start point to replicate its DNA, eukaryotic cells use multiple replication origins [Huberman and Riggs, 1968; Linskens and Huberman, 1990]. The existence of multiple replicons—chromosomal segments that are replicated from a single origin and whose size, number, and temporal order of replication is cell type– and developmental stage–specific [Edenberg and Huberman, 1975; Hand, 1978]—has prompted the idea of chromatin compartmentalization into domains.

Origins of replication specific for only a handful of genes have been identified in mammalian cells. They include those of the human c-*myc* gene [Iguchi-Ariga et al., 1988; Leffak and James, 1989; Ariga et al., 1989; Vassilev and Johnson, 1990], the human *HSP70* gene [Taira et al., 1994], the human choline acetyltransferase gene (Boulikas et al., in preparation), the Chinese hamster *DHFR* gene [Anachkova and Hamlin, 1988; Burhans et al., 1990; Caddle et al., 1990b; Vassilev et al., 1990], ribosomal protein *S14* [Tasheva and Roufa, 1994a,b] and rhodopsin [Gale et al., 1992] genes, the mouse adenosine deaminase locus [Carroll et al., 1993; Virta-Pearlman et al., 1993], and the mouse enhancer region of immunoglobulin heavy chain gene [Arizumi et al., 1993; Iguchi-Ariga et al., 1993].

An understanding of the function and sequence characteristics of the origin of replication is of utmost importance to our knowledge of the mechanisms underlying the differential expression of genes. Indeed, the origin of replication seems to play an important role in gene activity, and active genes seem to possess an actively used origin, usually in their 5′ flanking region, whereas inactive genes are thought to be replicated from an origin in their 3′ flanking region [Seidman et al., 1979; Taylor, 1984]. Thus, replication of the transcriptionally active c-*myc* and histone H5 genes occurs from origins lying in the 5′ flanking region, whereas these genes are replicated in the opposite direction from 3′ flanking origins in cell types harboring

Abbreviations used: ARS, autonomously replicating sequence; BPV, bovine papilloma virus; DHFR, dihydrofolate reductase; EBV, Epstein-Barr virus; HSV, herpes simplex virus; MAR, matrix-attached region; OBP, origin-binding protein; OBR, origin of bidirectional replication; ORI, origin of replication; RF-A, replication factor A; SV40, simian virus 40.

Received October 20, 1994; accepted December 19, 1994.

them into transcriptionally inactive conformations [Trempe et al., 1988; Leffak and James, 1989]. However, replication of the active *DHFR* locus in CHO cells occurs from a position in the 3′ flanking region [Tasheva and Roufa, 1994b].

About 50,000 origins of replication may be present in each nucleus in mammalian cells during all developmental stages; a significant fraction of these might be used at early and fewer at later stages of development [Spradling and Orr-Weaver, 1987]. A fraction of ORIs, linked with housekeeping genes, might be active at all stages of development. Thus, a decrease in the number of active ORIs takes place during development [Callan, 1974]. Models have been proposed showing that different origins of replication are inactivated or activated during embryogenesis as committed cell types evolve; this process was thought to play an important role in the expression of developmentally controlled genes from a coupling between replicational control and transcriptional competence [Smithies, 1982; Goldman et al., 1984; Taylor, 1984; Villarreal, 1991; Boulikas, 1995c]. For example, new ORIs might be activated in the β-globin gene complex that were not used in early embryo or fetal stages; such events of ORI activation might be connected to the developmental activation of the β- and the inactivation of the ϵ- and γ-globin genes [Boulikas, 1993b].

It has been established that active genes, found in decondensed chromatin structures, replicate during the early part of the S phase, whereas nontranscribed sequences confined into condensed chromatin structures, such as satellite DNA, replicate in late S [Remington and Flickinger, 1971; Balazs et al., 1974; Goldman et al., 1984; Taylor, 1984; Holmquist, 1987; Dhar et al., 1988]. Studies on the timing of replication of individual genes in gene clusters, such as the α- and β-globin, the immunoglobulin κ, the T-cell receptor β-chain, and the c-*myc* loci, have shown that all genes in a cluster replicate together; an early-replicating gene was never detected within a late-replicating cluster [Hatton et al., 1988]. The timing of replication of the various genes during the S phase could play a role in gene expression through the selective binding of DNA sequence-specific transcription factors to genes [Gottesfeld and Bloomer, 1982].

Apparently, mechanisms for triggering origins of replication at different time intervals of S phase and developmental stages have been evolved; DNA replication foci sequester almost all of the DNA methyltransferase activity [Leonhardt et al., 1992], and thus DNA methylation might be involved in differential activation of ORIs during S phase. Short segments in mammalian ORIs appear to become heavily methylated only in replicating but not in resting cells [Tasheva and Roufa, 1994b] (see level of DNA Methylation in ORIs). In addition phosphorylation of transcription and replication factors is a key regulator of the replication process [Boulikas, 1995a,b] and in regulating the triggering once and only once in each S phase of initiation from a single origin among thousands [Virshup, 1990; Huberman, 1991]. The dephosphorylated or phosphorylated initiator proteins would be inactivated and unable to initiate until the next cycle. In this model, distinct initiator proteins act upon different ORIs or groups of ORIs at S phase intervals.

In the present study we have examined most known origins of replication among all organisms for common structural/functional features; several rules diagnostic of origins of replication are deduced or proposed.

RESULTS AND DISCUSSION

Short Homo-Oligonucleotide Repeats in ORIs

The ORI sequences seem to possess an unusually high number of tri-, tetra-, and higher nucleotide repeats, more often dispersed rather than in tandem arrangements. For example, *oriC* of *E. coli* contains a number of GATTC, GATCC, AGATCT, and GATC motifs as well as their complementary sequences GAATC, GGATC, AGATCT, and GATC (shown underlined in Fig. 1). On pure probability considerations a tetranucleotide occurs once every 256 nucleotides; the GATC occurs 17 times in the 519 bp *E. coli* ORI, and it is thus overrepresented (8.4 times over random DNA); eight GATC motifs within a 245 bp sequence of *oriC*, representing the minimal origin of replication, are phylogenetically conserved [Meijer et al., 1979; Oka et al., 1980; reviewed by Zyskind and Smith, 1986]; this tetranucleotide is recognized by *dam* methyltransferase which methylates the adenine residue at the N^6 position to regulate initiation once and only once per cell cycle [Russell and Zinder, 1987; Boye and Løbner-Olesen, 1990].

A number of simple homo-oligonucleotides like CCC, GGG, TTT, and AAA are also overrepresented in *oriC* of *E. coli*. Such simple homo-oligonucleotide repetitions are found in most ORIs examined (see Figs 2–7); they might be

E.coli oriC (519 bp)

AGCCCGGGCCGTGGATTCTACTCAACTTTGTCGGCTTGAGAAAGACCTAGGA
TCCTGGG***TATTAAAAAGAAGATCTATTTATTT***AGAGATCTGTTCTATTGTG
ATCTCTTATTAGGATCGCACTGCCCTGTGGATAACAAGGATCCGGC***TTTTAA***
GATCAACAACCT**GGAAAGGA**TCATTAACTGTGAATGATCGGTGATCCTGGA
CCGTATAAGCT**GGGATCAGAATGAGGGG**TTATACACAACTCAAAAACTGA
ACAACAGTTGTTCTTTGGATAActaccggtTGATCCAAGCTTCCTGACAGAGT
TATCCACAGTAGATCGCACGATCTG***TATACTTATTT***GAG***TAAATTAA***CCCA
CGATCCCAGCCATTCTTCTGCCGGATCTTCCGGAATGTCGTGATCAAGAAT
GTTGAtcTCAGTGTTTCGCCTGTCTGTTTTGCACCGG***AATTTTT***GAGTTCT
GCCTCGAGTTTATCGATAGCCCCACAAAAAGGTGTCATATTCACGACTGCCA
ATAC

Fig. 1. *OriC* of *E. coli.* The single origin of replication of the *E. coli* genome is shown; the tetranucleotide GATC as well as the motifs AGATTC and its related AGATCC, AGATCT, and AGGATC and their compementary GAATCT, GGATCT, AGATCT, and GATCCT are shown underlined; also, simple GGG, TTT, CCC, and AAA repeats are present; binding of dnaA is essential for initiation of replication; the four dnaA sites TTAT(C/A)CA(C/A)A or its complementary T(G/T)TG(G/T)ATAA are double-underlined; AT-rich tracts are boldface italics; the DNA unwinding element consisting of three repeats of the sequence GATCTnTTnTTTT which is thermodynamically unstable in supercoiled DNA and is very sensitive to the single-strand-specific mung bean nuclease is shown dot-underlined [Kowalski and Eddy, 1989]; some eminent inverted repeats are outlined with the loop in lowercase; GA-rich tracts are in boldface. From Meijer et al. [1979] and Fuller et al. [1984].

remnants of evolution or simply products of extensive errors during replication of origins.

In support of this rule, the yeast replication factor MCM1 recognizes among others the sequence GATATTTCCAATTTGGGAAATTTCCCAAATCAGTAAT [Kuo and Grayhack, 1994]; the putative replication initiator protein RIP-60 in mammals recognizes tandem repeats of ATT [Dailey et al., 1990]. Sequences containing similar homo-oligonucleotide repeats in ORIs may constitute recognition sites of similar or unrelated to MCM1 and RIP-60 replication factors.

Initiator Protein Sites in Core ORI

Origin cores usually possess two to four repeated sequences of 5–20 bp each that represent binding sites of an initiator protein; binding of this initiator protein would distort the DNA causing initiation by local unwinding that is propagated to nearby AT-rich sequences.

Large T of SV40 binds to the four GAGGC sequences (Fig. 2A) at the 5,232, 5,236, 5, and 11 positions of the 5,243 bp genome [DeLucia et al., 1983; Tegtmeyer et al., 1983; Jones and Tjian, 1984; Deb et al., 1986a]. These sites which are part of a palindrome (outlined) form two pairs, **GAGGCCGAGGC,** arranged in opposite orientation; GAGGC repeats in each pair are 6 nt apart and the two pairs 7 nt apart; these T antigen sites are immediately flanked by an AT-rich region of 17 nt on the 3′ side (bold italicized in Fig. 2A) and by the early palindrome of 20 nt at the 5′ side [Deb et al., 1986a] (outlined in Fig. 2A). In the presence of ATP, 12 molecules of large T antigen are assembled in the form of two hexamers on the SV40 core ORI [Mastrangelo et al., 1989; Tsurimoto et al., 1989]. According to the model of Parsons and coworkers [1991], assembly of an hexamer first occurs on the early half core ORI and then on the late half; the formation of these hexamers melts the early and untwists the late half core ORIs; melting and untwisting release the large T antigen molecules from the GAGGC pentanucleotides to act as helicases at the flanking AT tract.

A single origin, *oriC,* is used by *E. coli* for the initiation of its replication. During rapid growth, an *E. coli* chromosome may acquire multiple copies of the chromosomal origin region. The timing of replication is controlled by the initiation-specific protein DnaA [Løbner-Olesen et al., 1989; reviewed by Georgopoulos, 1989] and the *dam* methyltransferase [Boye and Løbner-Olesen, 1990; Campbell and Kleckner, 1990]. About 30 monomers of dnaA protein in *E. coli* interact with *oriC.* This binding eventually causes structural alterations to the *oriC;* both DNase I cleavage and electron micrographs of the large dnaA-*oriC* complexes are consistent with wrapping of DNA around a large complex of dnaA molecules and melting of the DNA [Fuller et al., 1984; Bramhill and Kornberg, 1988].

A. SV40 ORI (205 bp)

5,146G***TTTAAAACTTTAT***CCATCTTTGCAAAgcttTTTGCAAAAGCCTAGGC
CTCCAAAAAAGCCTCTTCACTACTTCTGGAATAGCTCAGAGGCC
GAGGcggCCTCGGCCTCTGC***ATAAATAAAAAAAATTA***GTCAG
CCATGGGGCGGAGAATGGGCGGAACTGGGCGGAGTTAGGGGCGGGATGGG
CGGAGTTAGGGGCGGGACTAT107

B. λ ORI (185 bp)

GACC***AAATAAAAA***CATCTCAGAATGTGCATCCCTC**AAAAACGA**
GGGAAAA*TCCCCT***AAAACGAGGGATAAAA**CA***TCCCTC***AAATT**
GGGGGATTG***CTATCCCTC*****AAAACAGGGGGGACACAAAAGA**CA
C***TATTACAAAAGAAAAAAGAAAAGATTATT***CGTC**AGAGAA**TT
CTGGCGAA***TCCTCTC***GAC

C. Yeast *ARS1* (127 bp)

AATTTCGTC*AAAAATGCT*AAGAAATAggt***tatta***ctgagtag*T***
*ATTTATT*****TAA***GTATTGTTTGTGCACTTGCCTGCAGGCCTTTT
Gaaaagcaagc***a****TAAAAGATCT*AAAC***ATAAAAT***CTG***TAAAA***
T

D. EBV oriP core ORI (146 bp)

GATATCGCTG**TTCCTT**AGGA**CCCTTTT**ACTAACCCTAATTCGATAGCAT
ATGCTTCCCgttGGGTAACATATGCTATTGAATTAGGGTTAGTCTGGA
TAGTATATACTACTACCC**GGGAAG**CATATGCTACCCGTTTAGGGTTAAC

E. EBV replication enhancer; 30 bp repeat (repeated 20 times) (600 bp).

GGATAGCatatGCTACCCAGATATAGATTAGGATAGCatatGCTACCCA
GATATAGATTA

Fig. 2. **A:** The origin of replication of SV40 genome showing the four binding sites of T antigen within core ORI. SV40 genome is 5,243 bp; T antigen sites I, II, and III are double-underlined; CG-rich tracts (Sp1 sites) are underlined; core ORI (nucleotides 5,210 to 5,243/0 to 31) is in larger font; the two GAGGC and its complementary GCCTC sites within T antigen site II are double-underlined; a 12/12 palindrome in T antigen site II with CAGAGGCCGAGG in the stem is outlined; the CGG in the loop is in lowercase; the central G in the loop of the 12/12 palindrome is nucleotide 5,243/0; structural distortions in site II are transferred to the 17 bp AT-rich region in core origin (bold-face italics), causing its melting. From Subramanian et al. [1977] and Hay and DePamphilis [1982]. **B:** The ORI of bacteriophage λ. The palindromic O protein binding sites are double-underlined. From Hobom et al. [1979]. **C:** Yeast *ARS1* (127 bp). The ABF1 site is boldface and dot-underlined; the site of an unknown protein factor is underlined; the site of interaction of the origin recognition complex (ORC) determined by genomic footprinting is double-underlined; two inverted repeats are outlined. From Diffley and Cocker [1992] and Bell and Stillman [1992]. **D:** EBV oriP core ORI. The EBNA 1 sites are double-underlined; the 65 bp dyad symmetry containing three mismatches in core ORI is shown outlined with the three nucleotides in the loop in lowercase; the 960 bp of the unique sequence between the 30 bp repeats (enhancer) and the core ORI can be deleted without eliminating oriP activity. From Reisman et al. [1985], Frappier and O'Donnell [1992], and Hsieh et al. [1993]. **E:** EBV oriP replication enhancer. The replication enhancer is a 30 bp repeat repeated 20 times (600 bp); two repeats are shown; the 7 bp inverted repeat (one mismatch) is shown outlined with 4 nt in lowercase for the loop; evidently a cruciform of 18 nucleotides in each arm can be formed with two mismatches from two repeats and with the AGATATAGATTA central sequence in the loop; the EBNA1 sites are double-underlined. From Reisman et al. [1985], Hsieh et al. [1993], and Frappier and O'Donnell [1992].

Two λ proteins, O and P, are required for λ DNA replication; the O protein interacts with the core ORI causing structural distortion to the DNA double helix, best described by partial unwinding and wrapping of the DNA around an O protein complex [Dodson et al., 1986] (Fig. 2B). O protein recognizes, binds to, and bends the 19 bp sequence of hyphenated dyad symmetry [Zhan and Blattner, 1985]. P protein localizes the essential replication enzymes of *E. coli* at the *ori* λ site and by its simultaneous interaction with O protein and with DnaB generates a larger complex that includes additional DNA from the AT-rich region adjacent to the O binding site [Dodson et al., 1986].

Bell and Stillman [1992] described the purification from yeast cells of a six protein complex with molecular weights of 120, 72, 62, 56, 53, and 50 kDa; the origin recognition complex (ORC) binds with high sequence specificity to the *ARS1* yeast origin of replication that is AT-rich, causing local distortions of the double helix (Fig. 2C); this process requires ATP. The requirement of ATP for DNA binding of ORC, the DNase I cleavage pattern of DNA-ORC complexes compatible with wrapping of the 11 bp ARS and its 3′ flanking region around the ORC complex [Bell and Stillman, 1992], and the structural distortion caused to the DNA give to ORC three characteristics of bacterial initiator proteins [reviewed by Newlon, 1993].

In vivo footprinting using Raji cells has revealed that a single protein, EBNA1 (Epstein-Barr virus nuclear antigen 1), interacts with four repeats at the genetically defined replication origin of EBV, oriP, that includes the physical origin of bidirectional replication, OBR, and directs the cellular replication apparatus to initiate replication; these four sites contain the consensus EBNA1 recognition sequence GATAGCATATGCTACC [Hsieh et al., 1993] (Fig. 2D). A second cluster of 20 EBNA1 sites within oriP known as 30 bp repeats serves as a replication enhancer (Fig. 2E) and is located 960 bp away from OBR. The 30 bp repeats act in a cooperative manner independent of distance and orientation to activate EBV replication [Wysokenski and Yates, 1989] and also as a potent EBNA1-dependent transcriptional enhancer [Reisman et al., 1985; Sugden and Warren, 1989].

Interaction between EBNA1 molecules bound to the 30 bp repeats on one hand and to OBR on the other create a DNA loop, as observed by electron microscopy [Frappier and O'Donnell 1991b; Su et al., 1991]; this endows EBNA1 with properties characteristic of nuclear matrix proteins as, for example, SAF-A and RAP-1 [reviewed by Boulikas, 1995d]. This interaction might stabilize binding of EBNA1 to OBR, or it might attract the cellular replication machinery to OBR [Frappier and O'Donnell, 1991a]. Both the 30 bp repeats and the OBR sites are nucleosome free; EBNA1 was the only protein bound in these sites in vivo [Hsieh et al., 1993]. These studies are in agreement with in vitro studies showing that specific thymines in OBR but not in the 30 bp repeat are rendered sensitive to oxidation by permanganate as a result of distortion of the double helix caused by EBNA1 binding [Frappier and O'Donnell, 1992]. It can be speculated that the spacing of the four EBNA1 sites at OBR is such so as to cause more distortion to the double helix than its binding to the 30 bp repeats; pairs of the four EBNA1 sites at OBR are spaced at a center-to-center distance of 21 bp and the two pairs at a distance of 33 bp or of two and three helical turns, respectively [Hsieh et al., 1993] compared with the 1.5 helical turn spacing of EBNA1 sites at the 30 bp repeats.

UL9 initiator protein binds to three sites in the origin of HSV-1 ori_s (see Fig. 3A) as shown by DNase I and MNase footprinting [Koff et al., 1991]. This binding is cooperative [Elias et al., 1990]. It is thought that interaction of UL9 with sites I and II of the core ORI causes looping of the AT-rich stretch in between (Fig. 3A) [Koff et al., 1991]. $KMnO_4$ and DMS reactivity show that UL9 causes distortion at the AT-rich region; DNA supercoiling was absolutely required for distortion [Koff et al., 1991]. It appears that UL9 binds to sites I and II and loops the AT-rich region regardless of the phasing of the UL9 binding sites as shown by insertion of AT stretches within the AT tract. However, different tertiary structures are expected in the region of the looped DNA; the altered tertiary structure carrying an $(AT)_3$ insert might contribute to loss of replication function [Koff et al., 1991]. Interaction of UL9 with sites III might be transient, allowing for interaction of UL9 molecules bound to site III with UL9 molecules bound to sites I and II [Koff et al., 1991]; however, deletion or mutation in site III reduces both UL9 binding to sites I and II and DNA replication severalfold [Weir et al., 1989; Elias et al., 1990].

A. HSV-1 ORI$_S$ (104 bp)

540GGCCGCCGGGTAAAAGAAGTGAGAACGCGAAGCGTTCGCACTTCGTC
CCAATA***tatatatat***TATTAGGGCGAAGTGCGAGCACTGGCGCCGGCC
CCGGGATCC

B. Mouse rDNA ORI (muNTS 1) (421 bp)

ATCATGAAGGCACATTGG***ATTTTAT***GACAGAGTCTGTGTGTGT
GTGTGTG***TATAATATTT***CTGCTATGATTGCAGTTACTTGCCATC
TCGTGGGTTAGTTTGATTTCTGTAG***TTTTTTAAAATTCTTTAA***
AATTTTTATTTTATATTTTTTTAGTTTAGTTTAG***TTTAATTTA***
GTTTAGTTTTCAAGACAGGG**TTTCTCT**GTATAGCCCTGACTGTC
CTGGAACTCACTTTGCAGACCAGGCTGGCCTCAAACTCAGAAAT
CCTCCCATCTCTGCCT**GAAGAGAG**CTGGGATTAAAGACATGC
GCCATCACTCCCGGC***TATTTTTAAATTTTTAAATTATATTTAT***
TTAATTTATTTTTTTTGTTTTTTTCAAGATGTGGTTTCTCTGT
GTAAACTCTGGCTGACCGGAACTCACTGTGTA

Fig. 3. **A:** The origin of replication ORIs of HSV-1 showing the 18 bp AT-rich stretch between the binding sites of the initiator protein. HSV-1 (152 kb) contains one copy of ORI$_L$ and two copies of ORI$_S$; ORI$_S$ includes a 55 bp palindromic region (outlined) including sites I and II of UL9, highly homologous to the 144 bp palindromic region of ORI$_L$; the three sites (III, I, II sequentially from the 5′ to 3′ direction) of UL9, also known as OBP, defined by DNase I footprinting are double-underlined; GC tracts are underlined; AT tracts are in boldfaced italics; UL9 loops and distorts the ORI; the 11 bp site I to the left of the AT-rich region is the high affinity site also occurring twice in ORI$_L$; mutations in sites I, II, or III in ORI$_S$-containing plasmid DNA significantly decreases replication efficiency in transient replication assays. From Stow and McMonagle [1983], Elias et al. [1990, 1992], Weir et al. [1989], Weir and Stow [1990], Koff et al. [1991], and Dabrowski et al. [1994]. **B:** Mouse rDNA ORI (muNTS1). Putative ORI of replication and amplification of mouse rDNA (44 kb) in mouse L fibroblasts; the HMG I footprints are double-underlined; GA and CT tracts are in bold; AT tracts are boldface italics. From Wegner et al. [1989].

AT-Rich Stretches in ORIs

AT-rich regions of varying size are omnipresent components of origins of replication. AT-rich stretches might flank the core ORI, as, for example, in EBV ORI. Core ORI is defined as the short sequence ($\sim$50–100 nucleotides) where the initiator protein binds, causing local distortion in the double helix and initiating DNA unwinding. More often the AT-rich tract is included in the core ORI and may lie between two binding sites of the initiator protein, as is the case of the 18 bp AT stretch in the ORI$_S$ of HSV1 (Fig. 3A) [Elias et al., 1990]. AT-rich stretches are in bold italics in the figures of this work.

Five different functions can be assigned to AT-rich stretches.

First, the most conspicuous role of the AT-rich regions is that they facilitate DNA unwinding catalyzed by helicase molecules. A DNA unwinding element was described in the origin of *E. coli* as a GATCTnTTnTTTT element which is thermodynamically unstable, as evidenced by its sensitivity to the single-strand-specific mung bean nuclease [Kowalski and Eddy, 1989]. AT tracts serve as the points for initiation of DNA unwinding caused by the binding of the initiator proteins to the AT-flanking region; single-stranded DNA-binding proteins (as, for example, RPA) then interact with melted AT tract and attract the DNA polymerase α-primase complex.

Second, they might be involved in the attachment of the ORI to the nuclear matrix in eukaryotic cells where the majority of matrix-attached regions (MARs) appear to possess AT-rich tracts; a DNA unwinding matrix activity is found associated with rather short AT-rich motifs; indeed, the AATATATTT motif present within the MARs of both IgH gene and β-interferon genes becomes the nucleation site of a DNA unwinding effect under torsional strain that is propagated to neighboring sequences to include a 200 bp stretch of stably unpaired DNA [Dickinson et al., 1992].

Third, AT-rich stretches in ORIs might interact with HMG 1 and 2 as in the 50 bp AT-rich

stretches of the amplification origins located within the nontranscribed spacer of the murine rDNA [Wegner et al., 1989] (Fig. 3B); the rDNA repeats are 44 kb in size head-to-tail concatamers that probably resulted either from a rolling circle mode of replication of the initial rDNA plasmid or by multiplication and homologous recombination of already integrated plasmid DNA transfected into mouse L fibroblasts and integrated into the genome. According to this study, the AT-rich stretches located 4–5 kb upstream from the RNA polymerase I start site interact with the HMGI type of proteins, as revealed by footprinting, and function in ORI activation but not in transcription [Wegner et al., 1989].

Fourth, AT-rich stretches represent the binding sites of a special class of regulatory proteins; for example, yeast ARS elements include the (T/A)TTTA(T/C)(A/G)TTT(T/A) 11 bp sequence also found in ORIs in other species representing the binding site of the OBC protein complex (Fig. 2C) [Diffley and Cocker, 1992; Bell and Stillman, 1992]. It is proposed here that transcription factors may also function in replication. A recent compilation of DNA binding sites for transcription factors has revealed the presence of a class of proteins with an AT-rich sequence specificity [Boulikas, 1994a].

Fifth, although not the actual sites of binding of the initiator protein, AT tracts may be the principal site of local distortion of the double helix caused by the binding of the initiator protein to immediate flanking regions; this is the case for *E. coli oriC* [Bramhill and Kornberg, 1988; Kowalski and Eddy, 1989], λ bacteriophage ori [Schnos et al., 1988], SV40 ORI [Mastrangelo et al., 1989; Tsurimoto et al., 1989; Dean and Hurwitz, 1991], and yeast [Umek and Kowalski, 1988].

The AT-rich region might be part of a palindrome; this appears to be the case for the $(TA)_{16}$ element within the core origin of varicella-zoster virus (VZV) which forms the center of a potential cruciform structure of 46 bp as well as for the HSV $AA(TA)_5TTATTA$ element; as part of a palindrome, the AT tract may still represent the binding site of a transcription or replication factor.

Inverted Repeats in ORIs

Inverted repeats in ORIs have found their perfection in viruses: 144 bp with only two mismatches in HSV-1 ori_L (shown outlined in Fig. 4A and throughout this work with the central loop of the palindrome in lower case). Only about seven nucleotides in the stem are thought to be able to maintain a cruciform, and special computer programs are available to determine inverted repeats and their free energy of formation on DNA. Inverted repeats convert into cruciform structures when DNA is torsionally strained; initiation of both transcription and replication generates torsional strain. The role of a special class of inverted repeat-binding proteins and their function in stabilizing DNA in its cruciform structure is just emerging [Pearson et al., 1994; Xie and Boulikas, in preparation].

Inverted repeats may represent protein binding sites in the origin of replication. For example, cooperative binding of the UL9 protein (also called HSV-2 OBP) to palindromic sequences in HSV-2 or HSV-1 ori_s [Elias et al., 1990, 1992] causes looping of the AT tract located between them (Fig. 3A) [Koff et al., 1991]. The four binding sites of EBNA in oriP of EBV are short palindromic sites (Fig. 2D,E) [Frappier and O'Donnell, 1991a; Hsieh et al., 1993]. Similarly the binding sites of T antigen lie within an imperfect inverted repeat in SV40 ORI, shown outlined in Figure 2A [Deb et al., 1986b; Borowiec and Hurwitz, 1988; Parsons et al., 1990].

Inverted repeats are of two kinds: short (5–20 bp), usually representing the binding sites of initiator proteins, and long, best demonstrated by a 144 bp perfect inverted repeat in HSV-1 ori_L; long palindromes are believed to convert into cruciform structures and to act like sinks of torsional strain, thus facilitating unwinding of the double helix at the core origin. The presence of inverted repeats and AT-rich regions among monkey origins of replication has been noted before [Landry and Zannis-Hadjopoulos, 1991].

Several lines of evidence suggest cruciform formation at the time of activation of an origin of replication. First, monoclonal antibodies, directed against cruciforms occurring in ors sequences supposed to represent monkey origins of replication from unknown genes were shown to enhance DNA replication in permeabilized monkey cells [Zannis-Hadjopoulos et al., 1988]. This was interpreted to be a consequence of stabilization of the inverted repeats into their cruciform configuration by the antibody; such cruciforms were thought to be located at or near replication origins, resulting in multiple initiations of DNA replication from a single origin. Using monoclonal antibodies directed against

A. HSV-1 *ori* $_L$ (426 bp)

1ACCACGGGGTGCCGATGAACCCCGGCGGCTGGCAACGCGGGGTCCCTGC
GAGAGGCACAGATGCTTACGGTCAGGTGCTCCGGGCCGGGTGCGTCTGAT
ATGCGGTTGGTATATGTACACTTTACCTGGGGGCGTGCCGGACCGCCCCAG
CCCCTCCCACACCCCGCGCGTCATCAGCCGGTGGGCGTGGCCGCTA
TTATAAAAAAA.GTGAGAACGCGAAGCGTTCGCACTTTGTCCTAATAA
TA**tata**TATTATTAGGACAAAGTGCGAACGCTTCGCGTTCTCACTTTTT
TTATAATA.GCGGCCACGCCCACCGGCTACGTCACGCTCCTGTCGGCC
GCCGGCGGTCCATAAGCCCGGCCGGCCGGGCCGACGCGAATAAACCGGG
CCGCCGGCCGGGGCGCCGCGCAGCAGCTCGCCGCCCGG426

B. ori$_L$ of pseudo-rabies virus (628 bp)

CG**CCCCCTTTTTT**CGCGGCCCCC**GGGGAGCGGGGGG**TGGGGgccC
AGCTCTCCCCCGAGCGCGGATCTCTGAAAAAAAAATTTCC
CGCCCCGCGTTTTCCATTGGGGTGAATGGGGAGGGGG**G**
TGGGGGATGGGCACTCACTAAGAATTGGCAAGGTGCC***AATTTATATT***CTG
GCA**TGGTGCCAA**ACTACATACTGGCACCCTGCCAACCCCAA**TCCCCC**
TCCTACCT*TGGCACCCTGCCAACCCCAA* ***TCCCCCTCCTACC*** **CT**GGCAC
CCTGCCAACCCCAA**TCCCCC**ACCATCGTCGCGGTCTTCATCCTTGGCACA
CAGCCAACCCTAGAAGGCGAGCCGCGATGCTGCGGTCGGGCCCACCCGC
GGCAGGAGGTTGTCGATGGCGATGGGCGAGGGTGGGCGCGCGTGTGGTGG
GGACGGTGTCGGTGCCCGCGCCGGTGGGGCGAGGGTTCGCCTTGGGTGCG
CTGGGGGCCCCTCCCCATTCACCCCAATGGAAAACGCGGG
GCGGGAAATTTTTTTTTTCAGAGATCCGCGCTCGGGGGA
GAGCTGGGCCCCCGGTCAGCGAGTTGTGGGGGGGGGGGCTCGGGG
GAGACGGGCCCCCAGTCGAAGAGTTGTGGGGGCCCGGGG

Fig. 4. **A:** HSV-1 *ori*$_L$. The perfect 144 bp palindrome is shown outlined with the tata loop in lowercase; the three UL9 sites are double-underlined; GC tracts of more than six nucleotides are underlined; CT tracts are in bold. The HSV-1 genome is 160 kb and contains three origins; two copies of ORI$_S$ and one copy of ORI$_L$; ORI$_L$ is deleted during cloning in bacteria; its cloning was achieved using a yeast cloning vector; the 144 bp inverted repeat is similar to that of ORI$_S$. From Gray and Kaerner [1984] and Weller et al. [1985]. **B:** ori$_L$ of pseudorabies virus (628 bp). A perfect palindrome starting with TGGGGG in the first line and having 84 nucleotides in the stem and a very large loop of 353 nt (a size that can be occupied by two nucleosomes with a 20 bp linker) is shown outlined and in larger case; a site similar to UL9 of HSV-1 is in bold, double-underlined, and in larger case; an NFI consensus is double-underlined; a 34 bp CT-rich direct repeat present three times is shown dot-underlined with the second repeat in italics; the minimal ORI function was delimited to a fragment comprising only one arm of the palindrome but not the other nor the NFI and UL9 consensus sites; pseudorabies virus has a linear DNA that circularizes after penetrating the nucleus of the infected cell. From Kupershmidt et al. [1991].

cruciform and quantitative fluorescence flow cytometry, 3–5 × 10^5 cruciforms/nucleus were estimated for monkey CV-1 and human colon adenocarcinoma SW48 cells throughout S phase; however, no cruciform-like structures could be detected during G0 or G2M or in metaphase chromosomes [Ward et al., 1990, 1991]. Second, S1 nuclease sensitive sites appear as rodent cells move through G1 phase [Collins et al., 1982b]. The increase in the number of the single-stranded DNA-specific S1 nuclease cutting sites found during the S phase by Collins [1979] and Collins and coworkers [1982] support a model of a transient conversion of inverted repeats into cruciform structures during initiation of DNA replication. Third, formation of the two inverted repeats within the origin in the R1162 plasmid was found to be essential for activation of replication [Lin and Meyer, 1987].

The abundance of inverted repeats in a class of MARs [Boulikas and Kong, 1993a,b] believed to represent the sites of DNA replication also suggests a role of cruciform in origin function.

Intrinsically Curved DNA in Origins

A number of studies on origins of replication in viral and in other genomes sustain the idea that either the origin possesses intrinsically

curved DNA or that a severe bent is produced at the origin fragment as a result of its interaction with replication initiator proteins [Mukherjee et al., 1985; Zahn and Blattner, 1985; Koepsel and Khan, 1986; Deb et al., 1986b; Snyder et al., 1986; Schnos et al., 1988, 1989].

The fact that origins of replication coincide or colocalize with MARs and that MARs have been proven to possess intrinsically curved DNA from the retardation in mobility on agarose gels [see Boulikas, 1993a, and references therein] is one additional line of evidence for the presence of curved DNA in ORIs.

Many transcription factors are expected to prefer an intrinsically curved DNA binding site; this is in agreement with the view of the ORI/MAR elements as strong nucleation sites for transcription factors [Bidwell et al., 1993; Boulikas, 1994b] and as preferential sites of targeting by antineoplastic drugs [Boulikas, 1992b].

Classes of DNA Sequence Motifs in ORIs

Several classes of DNA sequence motifs have been found in ORIs:

1. The AT-rich class of motifs seems to constitute an omnipresent component of ORIs; AT-rich motifs (shown in boldfaced italics) are abundant in yeast ORIs (Fig. 5) and in the sequences flanking the heavily methylated 0.5 kb stretch of *DHFR* ORI (Fig. 6C).
2. A class of ORIs contains a large number of GA-rich tracts occasionally alternating with CT-rich tracts on the same strand of the DNA. Polypyrimidine or polypurine tracts (shown in boldface) are found in the λ ORI (Fig. 2B), in some stretches of the *RPS14* ORI (Fig. 6B), in the human c-*myc* ORI (Fig. 6A), and in the ORI of the human choline acetylransferase gene (not shown) (Boulikas et al., in preparation).
3. GC-rich tracts (shown underlined), as in the origin of the SV40 (Fig. 2A), HSV-1 ori_s (Fig. 3A), HSV-1 ori_L (Fig. 4A), and ori_L of pseudorabies virus (Fig. 4B).
4. The mixed motif is the major class of ORIs (Figs. 1–5A, 6, 7); however, mixed motif ORIs contain AT tracts and polypurines or GC-rich stretches.

Genomic Position of Origins

Most origins in multicellular organisms fall within upstream or downstream flanking regions of the gene they are linked with. For example, the ORI of the human c-*myc* gene (Fig. 6A) is at about −2 kb from the transcription initiation site, the ORI of the Chinese hamster *DHFR* gene (Fig. 6C) is 17 kb to the 3′ site, and that of the rDNA gene (Fig. 3B) is in the spacer region between two genes in the gene repeat.

However, as more studies are being completed, a more diverse picture with respect to the genomic position of the origins is emerging: the ORI of the S14 ribosomal protein gene in Chinese hamster cells comprises part of the coding and intronic regions (Fig. 6B) [Tacheva and Roufa, 1994a,b], and that of the human *HSP70* gene (−500 to −150) includes part of the promoter region [Taira et al., 1994]. The ORI of the human choline acetyltransferase gene includes the entire 5′ untranslated region as well as upstream flanking regions (Boulikas et al., in preparation). Initiation of DNA replication within a transcribed region has been found by Wu and coworkers [1993].

DNase I–Hypersensitive Sites in ORIs

ORIs are proposed here to constitute important control elements, colocalizing with transcriptional enhancers, tightly linked with the transcriptional activity and the temporal manner of replication of genes during the S phase intervals. Initiator proteins distort the DNA; it is proposed that ORIs become DNase I–hypersensitive at the time of their activation as a result of binding of transcription and replication DNA sequence-specific protein factors, as a result of unwinding by the nuclease sensitivity of the single-stranded region, and as a result of absence of nucleosomes that might need to be removed from core ORIs as a prelude to their activation [Hsieh et al., 1993]. Indeed, core ORIs have a local region of their double helix distorted at the time of activation that is rendered hypersensitive to DNase I and to some chemicals that attack single-stranded over double-stranded DNA such as $KMnO_4$ [Frappier and O'Donnell, 1992].

One additional reason to believe that core ORIs are into DNase I–hypersensitive conformations at the time of their activation emerges from studies in synchronized cells; it has been conjectured that cruciform structures appear at the origin only during their activation in S phase and that these structures disappear at all other stages of the cell cycle; the loop of the cruciform is into a DNase I–hypersensitive conformation [Ward et al., 1990, 1991].

A. Yeast C2G1 *ARS* (522 bp)

1GAATTCTAGGTGATATTGCAATTA**CTTCTTCTC**ATGCACTAACAAGTGAATG
ATAGAAATATGTTGAGTTGCTAACTGCCTG***ATTTTAAATAA***GTTTCATATTAT
AA**TCTTTT**AGC***ATATATATATATATATATT***GA**TCCTCTCTCTTCTTTATTTT**
CTGCCAGTAACCCATGTGT**GAAGAAG**AAAAC***ATAAATAAAAAA***GCAGTA
GCACATGGACACATTCACGCCCGAACACTTC***TAAAAA***GCAGCCCACAC**AA**
GAAAGTAG***ATATAAT***GTAGGACACCCAGCTTGTCC***ATAATT***GCTAATAGCA
TACTCAGGATAAC***ATATATTAAT***GACGACTCGTTTGCTCCAACTCACTCGTC
CTCATTACAG***ATTATTAT***CCCTA**CCTCTCC**AGAAA**CCCTTC*AATATAAAA***
AGGGCAGATGTCCGCTGCGAA**CCCTTCTCC**ATTTGGC***AATTATTT***GAACA
CCATCACTAAGTCCCTACAACAG***AATTTA***CAAACATG**CTTTCATTTCC**AAG
C**AAAAGAAA**TCGAT522

B. Yeast ARS element pARS772 (1044 bp)

gATcgTAcg**AAA**TAATcgTATAAgcAAAAggAAccATcTTTAcgTTcAAgTgTTTTT
TgcATTATTTgTcTTTcATTAAcTATTAcATgTcTATcAAcTgTAAcAAgTTTcAAA
AAggAAcgAgTAgAcATTTTAAATggTTAgccgTTATAAAATAcAATAcTATTccT
AgTAAAAATTTAAAggTTTTTTTcAcgTATcAcAAcTATcgTTgTgcgTATTTTATT
TTcTTATTTTAgTgAATcAAggTTgcAcATTATAcTAAAggTTTATTTATTTTTgggc
ATTTgTTTgAgAcAcgAATATTAATggcATAgccTTcAAATTggTTcgTggAAcATg
AATgTcAAATTTTTTAcTAgcATggcAAAgcATTATAAgAcTgTTgAATTATTTgTT
TAgTTATTAAATgATTAcccTcccAgATTTTAccggTcAATTcAAcAAAgAAAATAT
ATTTATATgAgATTAgATTTTTATAATTTATTTgTAAAAATgATATcgTTTcAgTT
gTTTTTggTTTTAATAAATTATTTgTTTATTTATTTATTTTTTTTTTTTTTAATTATA
ATATTTTATccAATccTTTTggTcAAAcAAATAAAcAgATTTAcTAcTTAgAAAAc
ggcATAAAAggAcAATAATAATAgAAcTATATTcAAAAcTgcAATTgATAAAAg
TAATTTTTAgTAAATATAAAAAgTATTgAAcAgcAgTTTgTTTcgTAgAccTAcTc
ATTATTcAcTgTAAATAgggAAgcAAcAAATTTTTTTTATgTgcTTTTTTAAAAAT
TTTAAATAcTAcgAAgcggcATATATTAAAAAAcTAAAATTATcTAcAAcAATA
AcAcAATAAcTcTTAgcTgTTcAATAgAgTccgAATTATTTcTgTAcATgTTTAAggc
gAATcAATTATTTATcATTTTcTgTcAAAAgATATcAAATgTTggTgcAcAATgcTA
gAAgATgTgTTATTAATTgATTTcTTTTTgAATgATAATTTTcTTAATTTTTATAAT
AgTTTATATgTgAAATAcAATTgATc

Fig. 5. **A:** Yeast C2G1 *ARS*. From chromosome III of *Saccharomyces cerevisiae;* ARS function can be provided by two copies of the ARS consensus (A/T)TTTA(A/G)TTT(A/T) (11 bp ARS core) shown double-underlined or one copy of the 11 bp ARS and additional sequences 3′ to the T-rich strand of the core; AT tracts are in boldface italics; CT and GA tracts of more than six nucleotides are in bold. From Palzkill and Newlon [1988]. **B:** Yeast *ARS* element pARS772. LOCUS SPARS772 PLN 14 MAY 1991. This ORI segment has 413 T (40%) and 365 A (35%) and a total of 75% AT and 25% GC. A is shown in bold and uppercase; T is in uppercase; G and C are in lowercase.

Highly Conserved and Hypervariable Motifs in ORIs

Origins of replication are proposed to be made up of highly conserved elements representing protein binding sites, surrounded by hypervariable sequence elements displaying variability not only across species but even among individuals of the same species. For example, most of the hypervariability in the mitochondrial genome is found at the origin of replication [Chang et al., 1985]; however, mitochondrial origins contain the highly conserved D-loop region (Fig. 7).

Thus, origins may be composed of short evolutionarily conserved DNA sequences that either include the inverted repeats or other protein-binding sites surrounded by highly variable regions displaying little sequence similarity between species.

The presence of curved DNA at the origin and the interaction of transcription and replication factors with the origin that distort the flanking regions which are then prone to small reactive molecules may constitute one mechanism proposed to contribute to hypervariability in ORIs.

ORIs are enriched in homo-oligonucleotide repeats (see Short Homo-Oligonucleotide Repeats in ORIs above); we do not know at present whether these elements are conserved and how they might have been derived during evolution. The presence of simple homo-oligonucleotide repeats in ORIs and of short direct repeats representing the binding sites of the initiator proteins (see Initiator Protein Sites in Core ORI above) are expected to generate slippage structures, thus causing small duplications during replication. The enrichment of ORIs in inverted repeats is also expected to produce errors during replication.

The presence of hypervariable regions flanking highly conserved elements within ORIs is indirectly supported from experimental data: a 0.5 kb stretch composed of a 30 bp AT repeat in the 3′ flanking region of the human apolipoprotein B gene is a MAR [reviewed by Boulikas, 1993a] and a hypervariable region [Boerwinkle et al., 1989]; although all MARs are not ORIs, it might be that an ORI includes or flanks this hypervariable region.

Level of DNA Methylation in ORIs

DNA methylation in *E. coli* determines origin activation. Only *E. coli* molecules with unmethylated A residues in GATC tetranucleotides within *oriC* replicate efficiently, whereas hemimethylated *oriC,* as is the newly replicated *oriC,* is not an efficient start point for DNA replication [Russell and Zinder, 1987; Boye and Løbner-Olesen, 1990]. Hemimethylated *oriC* is sequestered into the cell membrane, where it interacts with specific proteins [Hendrickson et al., 1982] and thus is unavailable for initiation [Ogden et al., 1988]. Experiments in which the cellular level of *dam* methyltransferase in *E. coli* cells was controlled by a temperature-inducible promoter have demonstrated the director's role of this enzyme in orchestrating the timing of initiation of DNA replication [Boye and Løbner-Olesen, 1990].

In line with experiments showing specific methylation in promoter, enhancer, and other regulatory regions as well as in coding regions of inactive but not of active genes in higher eukaryotes, one would expect DNA methylation at CpG to be higher in developmentally inactivated ORIs. A study on DNA methylation in CHO cells, however, using PCR and a genomic sequencing protocol able to pinpoint all 5-methylcytosines [Frommer et al., 1992] has shown that a defined 127 bp region within the ORI of the *RPS14* locus (Fig. 6B) is densely methylated at almost all C residues, including CpG, CpA, CpT, and CpC dinucleotides, only during replication [Tasheva and Roufa, 1994b]. Similar data were found for a 516 bp region within the ORI 17 kb to the 3′ side of the *DHFR* gene locus [Tasheva and Roufa, 1994b]. Such densely methylated islands could mediate the association of ORIs to the nuclear matrix, function in licensing particular ORIs to inactive states during execution of specific developmental programs, or be used by cells as signposts to mark previously replicated ORIs [Tasheva and Roufa, 1994b].

Developmentally Controlled ORIs May Be Enriched in ATTA and Related Homeodomain Protein Binding Sites

An enrichment of ATTA motifs representing recognition and binding sites of homeodomain proteins may be found in developmentally controled ORIs (shown in larger case in Figs. 6C, 7A) [Boulikas, 1992a]. A statistical treatment of this idea should await the identification of ORI stretches among the sequences of the GenBank. ATTA motifs may be more abundant in the AT tracts of ORIs and might contribute to the execution of the developmental program of silencing (and less often of activation) of specific ORIs during cell type formation; indeed, homeodomain proteins execute the body plan in all organisms studied, and mutations in their genes give severe disruptions in body parts [Scott et al., 1989]. In support of this hypothesis, RIP-60, a candidate replication initiation protein, with binding sites within the ORIs of the *DHFR* locus [Dailey et al., 1990], rhodopsin locus [Gale et al., 1992], and murine adenosine deaminase locus [Virta-Pearlman et al., 1993], recognizes the ATTATTATTATT or a repeat of the $(ATT)_n$ motif.

One Class of ORIs May Be Enriched in Potential Left-Handed and Triple-Helix-Forming Stretches

Several polypurine or polypyrimidine stretches possessing a mirror symmetry may be identified in ORIs; such sequences can form triple-helical structures under special conditions of torsional strain, pH, or salt [Johnston, 1988]. For example, the core ORI of bacteriophage λ has the stretch **AAAACGAGGGATAAAACATCCCT**-caaa**TTGGGGGATT**gcta**TCCCTCAAAAC AGGGGGACACAAAA** with the mirror symmetry motifs shown in bold and the center of the symmetry outlined. This motif is immediately followed by **TATTACAAAAGAAAAAGAA**-

AAGATTAT (see Fig. 2B) which also possesses a mirror symmetry.

A polypurine repeat of 248 bp is present within the ORI of mouse *ADA* locus; it contains stretches of potential triplex and slippage structures such as in the segment $A_5(GGAAA)_2$-(**GGAA**)$_9(AG)_4AAA(GGAA)_2GG(AG)_5AAAG$-$AAA(GGAA)_2GTAGA(GAAA)_2(GA)_3GGGAGA$-$(GGGA)_2G(GAAA)_2$ [Virta-Pearlman et al., 1993]. Evidently an extensive mirror symmetry motif is present in the 36 bp segment $(GGAA)_9$ shown in bold.

The ORI of the mouse rDNA gene repeat contains the stretch TTTTTTAAAATTCTTTAAAATTTT-TATTTTATATTTTTTTAGTTTAGTTTAGTTT-AATTTA145bpTATTTTTAAATTTTTAAATTA-TATTTATTTAATTTATTTTTTTTGTTTTTTT which includes two large footprints of HMG 1 separated by a nucleosome length of 145 bp (Fig. 3B). Mirror symmetries, inverted repeats, and slippage structures might be formed between or within these AT-rich sequences which can be brought together as the two linkers flanking the same nucleosome.

The mouse rDNA ORI contains the potential Z-DNA stretch $(TG)_8(TA)_2$ (dot-underlined in Fig. 3B). The *RPS14* ORI includes a $(CA)_7C$ motif (dot-underlined in Fig. 6B). Other more extensive potential Z-DNA elements may reside in the sequences flanking potential core ORIs [Boulikas, 1993b]. An examination of the 6.2 kb zone of replication of the amplified *DHFR* locus in CHO cells has indeed identified potential Z-DNA and triple helices that were also detected by their sensitivity to mung bean nuclease [Caddle et al., 1990b]. We are unable to perform a statistical treatment of Z-DNA stretches among ORI and non-ORI sequences because we cannot identify ORI sequences in the GenBank.

Z-DNA and triple-helical structures in origins are expected to be prone to mutagens and have been proposed to contribute to the low degree of conservation of some stretches in ORIs flanking elements highly conserved during evolution [Boulikas, 1992b].

RULES DIAGNOSTIC OF ORIS IN MULTICELLULAR ORGANISMS

Despite our incomplete knowledge of a statistically significant number of ORIs in multicellular organisms, we propose that most of the rules defined above are valid for origins in multicellular organisms. Two additional rules are deduced or proposed for ORIs in multicellular organisms.

Coincidence of MARs and ORIs

ORIs are in close proximity to or fall within the nuclear matrix attachment sites of chromatin loops. Indeed a number of studies rather conclusively demonstrate that initiation of DNA replication takes place on the nuclear matrix [reviewed by Boulikas, 1996d]. In addition, elongation of new DNA proceeds by reeling of the old strands through the matrix where the replication forks are anchored.

Excitement arose from the demonstration that a *Drosophila* SAR (i.e., an AT-rich region between the H1 and H3 genes of the 5 kb histone gene repeat) can function as ARS elements in both budding and fission yeasts [Amati and Gasser, 1990]. However, subsequent studies using BrdU heavy labeling and PCR aimed at mapping the sites of initiation of DNA replication in the histone gene repeat have challenged these results and shown the presence of multiple sites of initiation not coinciding with the SAR region [Shinomiya and Ina, 1993].

In support of this rule are studies on identification of mammalian ORIs showing the presence of consensus SAR sites such as in the ORI

Fig. 6. **A:** Human c-*myc* ORI. Located about 2.5 kb upstream of transcription initiation site, this sequence is both a transcriptional enhancer and an ORI; the autoregulatory TCTCTTA motif, binding site of c-Myc protein, is double-underlined; the remaining sequence proteced by c-Myc against DNase I is underlined. A footprint including the A_2T_9 motif is double-underlined [Saigo et al., 1994]. The FUSE (far upstream element) protein binding site at −1,551 [Avigan et al., 1990] that might contribute to ORI function is also shown double-underlined. The sequence is from Ariga et al. [1989]. **B:** Part of the Chinese hamster ribosomal protein *S14* gene ORI. Includes the third and fourth introns of the gene, the exons IV plus V, and about 500 bp of downstream flanking regions; a 127 bp region (larger case) is densely methylated (in every CpG, CpC, CpT, CpA) only in replicating cells. TG and GC motifs are underlined; GA- and CT-rich motifs are in boldface. From Tasheva and Roufa [1994a,b]. **C:** ORI of *DHFR* gene from Chinese hamster cells (position 2,431 to 3,550 of GenBank, accession number X52034, locus name CGDHFRORI). Region 2,807–3,322 (515 bp shown in uppercase) is heavily methylated only in replicating but not in resting cells. Transcription factor binding sites are double-underlined; these include, in the order of occurrence, RIP-60 (upstream RIP-60 site); TAACAGTAATAATAAATAT; Oct-1, ATTAGCAT; Oct-1, ATTTTCAT. ATTA and related motifs, putative core recognition sites of homeodomain proteins are in larger case. From Tasheva and Roufa [1994b], Caddle et al. [1990a], Held and Heintz [1992].

A. Human *c-myc* ORI (210 bp)

AAGCTTGTTTGGCCGTTTTAGGGTTTGTTGG***AATTTTTTTTT***CGTCTATGTAC
TTGTG***AATTATTT***CACGTTTGCCATTACCGGTTCTCCATAGGGTGATGTTCAT
TAGCAGTGGTGATAGG***TTAATTTT***CACATCTCTTATGCGGTTGAATAGTCAC
CTCTGAACC***AATTTTT***CCTCCAGTAA**CTCCTCTTTCTTC**GGA**CCTTCT**GC
AG(HindIII/-2328).....
(-1551 to -1527)CACAAAATAAAAAATCCCGAGGGAATATA(FUSE)

B. Part of the Chinese hamster ribosomal protein *S14* gene ORI (2kb)

AAGACATTGAGAGCGTTGATTTCTGTAAATG**CCCCTT**AACTG***TTATATTTT***
CCCATTTGCAATT**GGGAAGA**C***TTAAAAAA***CACTTGTGAA**TCTGTCCT**AGC
AG**CCTTTTTCCCT**GCTAGCTAAATGTT**AAGGG**TATG**CCCTTGGGA**CAGT
CCAAGTGAATAATTGCACACACCTGGCCTAAAGAGCAAACAGCCACAGGTG
CTCAGTGTAGGGGCTCTGAGTTAGAAACTT**GAAGGAGTCCTTCCCCTCT**G
ATGTTTGTGTG**CCTTCTC**AGAGCCTGTGTTGTCCTCGTCC**AATGTGGGTTG**
AGATGAGGAAATGACAGGTGAGGTGGGCCCTTGTTTATTGACCACTGG
TGATACATTT**GGGGAAGAATAAAG**TCTGCTAGTTAGTCTAATGGCT**AGGGA**
GGCTGTGTTTATGTAGCTATGGGCAATGACA**TTTTTTCTTCTT**GC**AGAGGA**
TGTCA**CCCCCATCCCCTCT**GACAGCACCC**GAAGGAAGGG**
TGGTCGTCGTGGTCGCCGTCTGTGAACAGGA**CT**
TCTC*AAATTATTTT*CTG***TTAATAAATT***GCTTTGT
ATAAGCTATTTTGGTTCTGATGTTTGTTTGTTTT**GA**
GGCAGGAT**CTCTCTCC**ATGTCTATGT**AGAAGAGTCTGGCCTCC**A
ACTCAGATCCACATA**CCTGTCTCT**G**CCTCCC**AAGTACT**GGGGATAAAGG**
AATGCACCACCACACCCCGGTGACCTTAGGGGCCACCCAGGGCAACACCA
AGGCAGTGTTTT**GAGGGGA**TGTTAAGATTGCATGTAGGAGCTTCTAATGTAG
GTTGGGGGTTGGCAAACCTGGGTTTGCAGGCTGGCTTTCT**GAGAATGGGA**
ACCT**GAGAGAGGAGA**TGCTGAGGCTGCCACTCTGGC***ATTAATTT***GCCAG
CCAG***TATTTTCAAA*****GTGGAA**CTGTTAGTAGCCCATGCTGTCATGGGACAG
CTAAACACCACACACACACACACCCCA

C. ORI of *DHFR* gene from Chinese hamster cells

ctcggcctgtctgtg***taat****atttaaaatgaaaa*ctttggaaatgttctgaaaccagctggtg
tcagatagtcagagaactttcgtaaggtaggtgtggGTTAtagcataatcccacacaagag
ct**gaagcaggagga**ttttgtgtttgagggcagctagagcacatggtgag**tccctgcct**ca
aaacac**aaaagcaagacaaaaacaag**ctccaaataagattcacttggg**cctttctttc**
cttccttctcagtgagtccacttc***ttttaaaat***caggtcttaaagacgcacttagatctgaatt
aacag***taataat****aaatat*cttctcttacagtacagattatgctctataaacactgcact
gataaagttcagCCTTAACCTTTGTTCTGTAAATGTTTCCTAGTTTTTCTACTG
CCGTATTATAAGACAAATGTCAGCAT**GAAGGCAGG**TTTTTCAGAAAACAC
AGCAGCTCACAGATGCCTTAATCATAATCATTAAAGCAGCTGCAACT
TTTTCAACT**GGGAAA**TCATTCAAGGATGT**TTTTT**CTGAAG**TCCCT**ACCAGGG
CACACGCACCTGGGTTGCTGTGTGACATCAGTTAGGTAGACTCTGAACTGG**C**
TTCCCAAGCAAATTATACA**AAAAGCAAGG**TGTCACCTAGTATTAGCAT
AACTTCTGATAACTACTGTCTTAGCTGGGGTTTCTATTGCTGT**GAAGAGA**
CACATGACCACAGAAACTC***TTATAAA***G**GAAAGC***A******ATTATT***GGGTCCA
GCTTACAGTTCAGAGGTTTAATCCATTGTCATGATTGC**AGGAAG**TA
TGGTGGCCCACAGCGAGACATGGTGCT**GGAGAAG**T**AGATGAGAG**TTCTAT
ATCAG**ATTG**ACACA**TTCTTCC**AACAAGGCCACACtcatcacttgagctatgggcc
attttcattcaaaccaccaaagctacaaggtagcttataccccagcttgctatttctgatgagact
tagtaaatagtc***ttaaaa***gccc***ataaaat***gactcaaaactag***ttttttattattattat***
tagttcaaatta**g**gaagaagcttgctttacatgtcaa**tcccttctccctctccctc**atcaaa
actag***ttttttgttttta***

Figure 6.

ORI_H of human mtDNA (612 bp).

GATCACAGGTCTATCACCCTATTAACCACTCACGGGAGCTCTCCATGCAT
TTGGTATTTTCGTCT**GGGGGG**TATGCACGCGATAGCATTGCGAGACGCTGG
AGCCGGAGCACCCTATGTCGCAGTATCTGTCTTTGATTCCTGCCTCATCC*T*
ATTA*TTTAT*CGCACCTACGTTCAATATTACAGGCGAACATACTTACT
AAAGTGTGTTAATTAATTAATGCTTGTAGGACATAATAATAACA
ATTGAATGTCTGCACAGCCACTTTCCACACAGACATCATAAC***AAAAAATTT***C
CACCAAA**CCCCCCCTCCCCC**GCTTCTGGCCACAGCACTTAAACACATCT
CTGCCAAACCCCAAAAACAAAGAACCCTAACACCAGCCTAACCAG***ATTTCA***
AATTTTATCTTTTGGCGGTATGCACTTTTAACAGTCA**CCCCCC**AACTAACA
C***ATTA****TTTT***CCCCTCCCACTCCCATACTACTAATCTCATCAATACAA
CCCCCGCCCATCCTACCCAGCACACACACACCGCTGCTAACCCCATACCC
CGAACCAACCAAACCCCAAAGACA**CCCCCC**ACAGTTTATGTAGCTTA**CCT**
CCTCAAAGCAATACACTGA

Fig. 7. ORI_H of human mtDNA. D-loop is 1–191; OBR is at nucleotide 191 (outlined) ATTA motifs are in larger case. From Anderson et al. [1981] and Chang et al. [1985].

of the Chinese hamster rhodopsin locus [Gale et al., 1992].

ORIs Are Activated by Transcription Factors

Transcriptional enhancers were proposed to act as replication enhancers to increase initiation of replication from the core origin; replication enhancers are to be found at great distance from the core ORIs. Thus, the functional origin of replication in our model is composed of the core origin and one or more replication enhancers at variable distances from the core origin [Boulikas, 1996c].

Due to their higher complexity over viral, bacterial, mitochondrial, and yeast ORIs, ORIs in multicellular organisms are proposed to possess binding sites for a higher number of protein transcription/replication factors in addition to the binding sites for the replication-specific initiator protein. This level of complexity allows their programming during embryogenesis and their differential replication during S phase intervals and is tightly coupled to gene expression. The tight coupling between replication and transcription might arise from the fact that most replication factors interacting with the core ORI and enhancer are transcription factors as well.

A glimpse of the complexity of eukaryotic origins arises from studies on ARS activation in yeast. This unicellular low eukaryote does not have to cope with the complexity of multicellular organisms that need to develop a specific program of silencing specific sets of ORIs during embryogenesis; nevertheless, its ORIs appear to be differentially activated during S-phase intervals [Ferguson et al., 1991]. The origin recognition complex (ORC) is a six protein complex that binds to the yeast ARS in vitro, causing a structural distortion at the double helix determined by DNAse I and copper phenanthroline [Diffley and Cocker, 1992]; this process seems to be more complex when compared with the structural distortion at the core ORI caused by a single protein such as the O protein or the dnaA protein in *E. coli,* the large T antigen in SV40 and polyoma, the EBNA1 protein in EBV, and the origin binding protein (OBP) in HSV-1 (Initiation Protein Sites in Core ORI).

Transcription factors implicated in replication

Animal viruses. A single protein may simultaneously regulate two entirely different processes: transcription and replication. Single proteins such as large T antigen and AP-1 have strong binding sites in the origin region and stimulate both replication and transcription [e.g., Hay and DePamphilis, 1982; Hendrickson et al., 1987; Martin et al., 1988; Murakami et al., 1991]. Sp1 stimulates SV40 DNA replication [Guo and DePamphilis, 1992]. The transcription factor NF-I (also called CTF, CCAAT box-binding protein, or C/EBP) stimulates replication of adenovirus DNA in vitro [Pruijn et al., 1986; Jones et al., 1987; Santoro et al., 1988; Coenjaerts et al., 1991] by establishing cooperative interactions with DBP (DNA-binding protein) of adenovirus [Cleat and Hay, 1989]. NF-I also stimulates replication in SV40 [Cheng and Kelly, 1989]; the

CCAAT element recognized by NF-I is an important promoter element for a significant number of eukaryotic genes [reviewed by Boulikas, 1994a]. The transcription factor Oct-1 (also called OTF-1 or NFIII) involved in the regulation of the histone H2B and immunoglobulin genes can stimulate initiation of adenovirus DNA replication in vitro [O'Neill et al., 1988; Mul et al., 1990; Verrijzer et al., 1990; Coenjaerts et al., 1991; reviewed by Challberg and Kelly, 1989]. The constellation of transcription factors implicated in the control of DNA replication may also include p53, a sequence-specific DNA-binding factor with a GC-rich sequence preference that might interact with the Sp1 site of SV40 ORI [Boulikas, 1994a]. The DNA sequence-specific activity RIP-60 might be a candidate replication initiator protein that binds to the *DHFR* replication origin [Caddle et al., 1990a; Dailey et al., 1990].

Yeast. The transcriptional activator GAL4 from yeast was capable of stimulating polyoma virus DNA replication; synergistic activation of DNA replication was achieved by multimerization of the GAL4 binding site [Bennett-Cook and Hassell, 1991; Baru et al., 1991]. The yeast protein ABF1 may contribute either to repression or activation of transcription, and its binding to several ARS elements may stimulate initiation of replication [Buchman et al., 1988; Diffley and Stillman, 1988]. The yeast protein MCM1 is a transcription activator that also affects DNA replication [Passmore et al., 1988; Christ and Tye, 1991]. Another cellular DNA binding protein that displays the property of regulating the initiation not only of transcription but also of replication is the RAP1/GRF-I factor in yeast [Brand et al., 1987; Kimmerly et al., 1988]. The origin recognition complex in yeast functions both in ORI firing and in transcription repression [Bell et al., 1993].

A Close-Up of Mammalian ORIs

One of the functions of c-Myc protein is to promote cellular DNA replication by binding to a cloned human putative ORI sequence [Iguchi-Ariga et al., 1987a]. c-Myc can even substitute for SV40 large T antigen in an in vitro SV40 replication system [Iguchi-Ariga et al., 1987b; Classon et al., 1987]. A region located approximately 2 kb upstream of the transcription start site of the human c-*myc* gene contains a putative ORI of 210 bp (Fig. 6A) which is also a transcriptional enhancer containing c-Myc binding sites [Iguchi-Ariga et al., 1988]. This fragment contains the TCTCTTATGCCGGTTGAATAGT Myc-binding site determined by DNase I footprinting and mobility shift assays; this interaction is involved in both upregulating transcription as well as replication of the c-*myc* gene domain [Ariga et al., 1989].

It is not clear at present whether or not c-Myc plays the role of the initiator protein; the fact the c-Myc protein can substitute for SV40 large T antigen in an SV40 DNA replication system in vitro [Iguchi-Ariga et al., 1987b] suggests that c-Myc is an initiator protein causing firing the origin of replication of the c-*myc* gene probably in concerted interaction with other transcription factors. The presence of three shift bands in mobility retardation assays [Ariga et al., 1989] suggest that c-Myc is cross-complexed with other transcription factors on this genomic DNA fragment. An activity was shown with DNase I footprinting to protect a region that includes the AATTTTTTTTT element of *c-myc* ORI [Saigo et al., 1994].

The 210 bp origin of the human c-*myc* oncogene (Fig. 6A) has the ability to act as an autonomously replicating sequence [Iguchi-Ariga et al., 1987a]. It contains three ATTA motifs; one ATTA motif is expected every 4^4 or 256 nucleotides by pure probability. The c-*myc* ORI contains a stretch of A_2T_9, a stretch of $A_2T_2AT_3$, a stretch of $T_2A_2T_4$, and a stretch of A_2T_5; the last two of these are in the almost immediate region flanking the c-Myc binding site (Fig. 6A). This fragment plays the role of a transcriptional enhancer for the expression of the c-*myc* gene [Iguchi-Ariga et al., 1988; Ariga et al., 1989].

DNA sequences enriched in origins of replication termed *ors* have been isolated by extrusion of single-stranded newly synthesized DNA at the replication fork from actively replicating monkey cells in culture [Zannis-Hadjopoulos et al., 1985]. pBR322 plasmid harboring several cloned *ors* sequences has been shown to be autonomously and extrachromosomally replicating after transfection into HeLa cells [Frappier and Zannis-Hadjopoulos, 1987; Rao et al., 1990; Landry and Zannis-Hadjopoulos, 1991]. These studies provide evidence that short sequences in the mammalian genome might represent core origins of replication.

The murine adenosine deaminase ORI (3,993 bp) contains a 377 bp polypurine stretch and binding sites for several transcription factors including the putative initiators PUR and RIP-60

and EBNA1 as well as Oct-1, C/EBP, and p53 implicated in stimulating DNA replication [Virta-Pearlman et al., 1993]. The DNA sequence-specific activity RIP-60 was inferred to be a candidate replication initiator protein that binds to the *DHFR* replication origin [Caddle et al., 1990a; Dailey et al., 1990]; RIP-60 copurifies with an ATP-dependent helicase and might aid the interaction of dispersed elements by DNA bending.

CONCLUSIONS

Thus, it seems that among all terrestrial life forms best studied in bacteria, bacteriophages, yeasts, and animal viruses, the activation of the origin of replication shares common fundamental characteristics. First, an initiator protein, such as dnaA in *E. coli,* O protein in λ bacteriophage, T antigen in SV40 and polyoma, and EBNA1 in Epstein-Barr virus, directs the formation of a large complex containing several initiator protein molecules. Second, the initiator protein complex causes local distortion at a small region of the core ORI and attracts other proteins of the DNA replication apparatus such as helicases and single-strand DNA-binding proteins. Third, an AT-rich DNA sequence of about 10–30 bp found in all core origins and flanking the initiator protein site melts. Fourth, the primary replication bubble at the AT-rich region is extended by helicases. This might turn out to be an universal mechanism exploited by higher eukaryotes as well for origin activation.

Evidently, knowledge of a large number of origins of replication could be used to control proliferation in cancer cells by oligonucleotide or other targeting methods.

REFERENCES

Amati B, Gasser SM (1990): *Drosophila* scaffold-attached regions bind nuclear scaffolds and can function as ARS elements in both budding and fission yeasts. Mol Cell Biol 10:5442–5454.

Anachkova B, Hamlin JL (1988): Replication in the amplified dihydrofolate reductase domain in CHO cells may initiate at two distinct sites, one of which is a repetitive sequence element. Mol Cel Biol 9:532–540.

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981): Sequence and organization of the human mitochondrial genome. Nature 290:457–465.

Ariga, H, Imamura Y, Iguchi-Ariga SMM (1989): DNA replication origin and transcriptional enhancer in c-*myc* gene share the c-*myc* protein binding sequences. EMBO J 8:4273–4279.

Ariizumi K, Wang Z, Tucker PW (1993): Immunoglobulin heavy chain enhancer is located near or in an initiation zone of chromosomal DNA replication. Proc Natl Acad Sci USA 90:3695–3699.

Avigan MI, Strober B, Levens D (1990): A far upstream element stimulates c-*myc* expression in undifferentiated leukemia cells. J Biol Chem 265:18538–18545.

Balazs I, Brown EH, Schildkraut CL (1974): The temporal order of replication of some DNA cistrons. Cold Spring Harbor Symp Quant Biol 38:239–245.

Baru M, Shlissel M, Manor H (1991): The yeast GAL4 protein transactivates the polyomavirus origin of DNA replication in mouse cells. J Virol 65:3496–3503.

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

Bell SP, Kobayashi R, Stillman B (1993): Yeast origin recognition complex functions in transcription silencing and DNA replication. Science 262:1844–1849.

Bennett-Cook ER, Hassell JA (1991): Activation of polyomavirus DNA replication by yeast GAL4 is dependent on its transcriptional activation domains. EMBO J 10:959–969.

Bidwell JP, van Wijnen AJ, Fey EG, Dworetzky S, Penman S, Stein JL, Lian JB, Stein GS (1993): Osteocalcin gene promoter-binding factors are tissue-specific nuclear matrix components. Proc Natl Acad Sci USA 90:3162–3166.

Boerwinkle E, Xiong W, Fourest E, Chan L (1989): Rapid typing of tandemly repeated hypervariable loci by the polymerase chain reaction: Application to the apolipoprotein B 3′ hypervariable region. Proc Natl Acad Sci USA 86:212–216.

Borowiec JA, Hurwitz J (1988): Localized melting and structural changes in the SV40 origin of replication induced by T-antigen. EMBO J 7:3149–3158.

Boulikas T (1992a): Homeotic protein binding sites, origins of replication and nuclear matrix anchorage sites share the ATTA and ATTTA motifs. J Cell Biochem 50:111–123.

Boulikas T (1992b): Chromatin and nuclear matrix in development and in carcinogenesis: A theory. Int J Oncol 1:357–372.

Boulikas T (1993a): Nature of DNA sequences at the attachment regions of genes to the nuclear matrix. J Cell Biochem 52:14–22.

Boulikas T (1993b): Homeodomain protein binding sites, inverted repeats, and nuclear matrix attachment regions along the human β-globin gene complex. J Cell Biochem 52:23–36.

Boulikas T (1994a): A compilation and classification of DNA binding sites for protein transcription factors from vertebrates. Crit Rev Euk Gene Expression 4:117–321.

Boulikas T (1994b): Transcription factor binding sites in the matrix attachment region of the chicken α-globin gene. J Cell Biochem 55:513–529.

Boulikas T (1995a): Phosphorylation of transcription factors and the control of the cell cycle. Crit Rev Euk Gene Expression 5:1–77.

Boulikas T (1995b): Control of DNA replication by protein phosphorylation. Anticancer Res. 14:2465–2472.

Boulikas T (1995c): A model of transcription/replication enhancers. Oncol Rep. 2:171–181.

Boulikas T (in press d): Chromatin domains and prediction of MAR sequences. Int Re Cytol. In press. Int Rev Cytol.

Boulikas T, Kong CF (1993a): A novel class of matrix attached regions (MARs) identified by random cloning and their implications in differentiation and carcinogenesis. Int J Oncol 2:325–330.

Boulikas T, Kong CF (1993b): Multitude of inverted repeats characterize a class of anchorage sites of chromatin loops to the nuclear matrix. J Cell Biochem 53:1–12.

Boye E, Løbner-Olesen A (1990): The role of *dam* methyltransferase in the control of DNA replication in E. coli. Cell 62:981–989.

Bramhill D, Kornberg A (1988): Duplex opening by dnaA protein at novel sequences in initiation of replication at the origin of the E. coli chromosome. Cell 52:743–755.

Brand AH, Micklem G, Nasmyth K (1987): A yeast silencer contains sequences that can promote autonomous plasmid replication and transcriptional activation. Cell 51:709–719.

Buchman AR, Kimmerly WJ, Rine J, Kornberg RD (1988): Two DNA-binding factors recognize specific sequences at silencers, upstream activating sequences, autonomously replicating sequences, and telomeres in *Saccaromyces cerevisiae*. Mol Cell Biol 8:210–225.

Burhans WC, Vassilev LT, Caddle MS, Heintz NH, DePamphilis ML (1990): Identification of an origin of bidirectional DNA replication in mammalian chromosomes. Cell 62:955–965.

Caddle MS, Dailey L, Heintz NH (1990a): RIP60, a mammalian origin-binding protein, enhances DNA bending near the dihydrofolate reductase origin of replication. Mol Cell Biol 10:6236–6243.

Caddle MS, Lussier RH, Heintz NH (1990b): Intramolecular DNA triplexes, bent DNA and DNA unwinding elements in the initiation region of an amplified dihydrofolate reductase replicon. J Mol Biol 211:19–33.

Callan HG (1974): DNA replication in the chromosomes of eukaryotes. Cold Spring Harbor Symp Quant Biol 38:195–203.

Campbell JL, Kleckner N (1990): E. coli *oriC* and the *dnaA* gene promoter are sequestered from *dam* methyltransferase following the passage of the chromosomal replication fork. Cell 62:967–979.

Carroll SM, DeRose ML, Kolman JL, Nonet GH, Kelly RE, Wahl GM (1993): Localization of a bidirectional origin in the native locus and in episomally amplified murine adenosine deaminase loci. Mol Cell Biol 13:2971–2981.

Challberg MD, Kelly TJ (1989): Animal virus DNA replication. Annu Rev Biochem 58:671–717.

Chang DD, Hauswirth WW, Clayton DA (1985): Replication priming and transcription initiate from precisely the same site in mouse mitochondrial DNA. EMBO J 4:1559–1567.

Cheng L, Kelly TJ (1989): Transcriptional activator nuclear factor I stimulates the replication of SV40 minichromosomes in vivo and in vitro. Cell 59:541–551.

Christ C, Tye B-K (1991): Functional domains of the yeast transcription/replication factor MCM1. Genes Dev 5:751–763.

Classon M, Henriksson M, Sümegi J, Klein G, Hammaskjöld ML (1987): Elevated *c-myc* expression facilitates the replication of SV40 DNA in human lymphoma cells. Nature 330:272–274.

Cleat PH, Hay RT (1989): Co-operative interactions between NFI and the adenovirus DNA binding protein at the adenovirus origin of replication. EMBO J 8:1841–1848.

Coenjaerts FEJ, de Vries E, Pruijn GJM, van Driel W, Bloemers SM, van der Lugt MT, van der Vliet PC (1991): Enhancement of DNA replication by transcription factors NFI and NFIII/Oct-1 depends critically on the positions of their binding sites in the adenovirus origin of replication. Biochim Biophys Acta 1090:61–69.

Collins JM (1979): Transient structure of replicative DNA in normal and transformed human fibroblasts. J Biol Chem 254:10167–10172.

Collins JM, Glock MS, Chu AK (1982): Nuclease S_1 sensitive sites in parental deoxyribonucleic acid of cold- and temperature-sensitive mammalian cells. Biochemistry 21:3414–3419.

Dabrowski CE, Carmillo PJ, Schaffer PA (1994): Cellular protein interactions with herpes simplex virus type 1 oriS. Mol Cell Biol 14:2545–2555.

Dailey L, Caddle MS, Heintz N, Heintz NH (1990): Purification of RIP60 and RIP100, mammalian proteins with origin-specific DNA-binding and ATP-dependent DNA helicase activities. Mol Cell Biol 10:6225–6235.

Dean FB, Hurwitz J (1991): Simian virus 40 large T antigen untwists DNA at the origin of DNA replication. J Biol Chem 266:5062–5071.

Deb S, DeLucia AL, Baur C-P, Koff A, Tegtmeyer P (1986a): Domain structure of the simian virus 40 core origin of replication. Mol Cell Biol 6:1663–1670.

Deb S, DeLucia AL, Koff A, Tsue S, Tegtmeyer P (1986b): The adenine-thymine domain of the simian virus 40 core origin directs DNA bending and coordinately regulates DNA replication. Mol Cell Biol 6:4578–4584.

DeLucia AL, Lewton BA, Tijian R, Tegtmeyer P (1983): Topography of simian virus 40 a protein-DNA complexes: Arrangement of pentanucleotide interaction sites at the origin of replication. J Virol 46:143–150.

Dhar V, Mager D, Iqbal A, Schildkraut CL (1988): The coordinate replication of the human β-globin gene domain reflects its transcriptional activity and nuclease hypersensitivity. Mol Cell Biol 8:4958–4965.

Dickinson LA, Joh T, Kohwi Y, Kohwi-Shigematsu (1992): A tissue-specific MAR/SAR DNA-binding protein with unusual binding site recognition. Cell 70:631–645.

Diffley JFX, Cocker JH (1992): Protein-DNA interactions at a yeast replication origin. Nature 357:169–172.

Diffley JFX, Stillman B (1988): Purification of a yeast protein that binds to origins of DNA replication and a transcriptional silencer. Proc Natl Acad Sci USA 85:2120–2124.

Dodson M, Echols H, Wickner S, Alfano C, Mensa-Wilmot K, Gomes B, LeBowitz J, Roberts JD, McMacken R (1986): Specialized nucleoprotein structures at the origin of replication of bacteriophage λ: localized unwinding of duplex DNA by a six-protein reaction. Proc Natl Acad Sci USA 83:7638–7642.

Edenberg HJ, Huberman JA (1975): Eukaryotic chromosome replication. Annu Rev Genet 9:245–284.

Elias P, Gustafsson CM, Hammarsten O (1990): The origin binding protein of herpes simplex virus 1 binds cooperatively to the viral origin of replication ori_s. J Biol Chem 265:17167–17173.

Elias P, Gustafsson CM, Hammarsten O, Stow ND (1992): Structural elements required for the cooperative binding of the herpes simplex virus origin binding protein to oriS reside in the N-terminal part of the protein. J Biol Chem 267:17424–17429.

Ferguson BM, Brewer BJ, Reynolds AE, Fangman WL (1991): A yeast origin of replication is activated late in S phase. Cell 65:507–515.

Frappier L, O'Donnell M (1991a): Overproduction, purification, and characterization of EBNA1, the origin binding protein of Epstein-Barr virus. J Biol Chem 266:7819–7826.

Frappier L, O'Donnell M (1991b): Epstein-Barr nuclear antigen 1 mediates a DNA loop within the latent replication origin of Epstein-Barr virus. Proc Natl Acad Sci USA 88:10875–10879.

Frappier L, O'Donnell M (1992): EBNA1 distorts *oriP*, the Epstein-Barr virus latent replication origin. J Virol 66: 1786–1790.

Frappier L, Zannis-Hadjopoulos M (1987): Autonomous replication of plasmids bearing monkey DNA origin-enriched sequences. Proc Natl Acad Sci USA 84:6668–6672.

Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL (1992): A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci USA 89:1827–1831.

Fuller RS, Funnell BE, Kornberg A (1984): The dnaA protein complex with the E. coli chromosomal replication origin (oriC) and other DNA sites. Cell 38:889–900.

Gale JM, Tobey RA, D'Anna JA (1992): Localization and DNA sequence of a replication origin in the rhodopsin gene locus of Chinese hamster cells. J Mol Biol 224:343–358.

Georgopoulos C (1989): The *E. coli* dnaA initiation protein: A protein for all seasons. Trends Genet 5:319–321.

Goldman MA, Holmquist GP, Gray MC, Caston LA, Nag A (1984): Replication timing of genes and middle repetitive sequences. Science 224:686–692.

Gottesfeld J, Bloomer LS (1982): Assembly of transcriptionally active 5S RNA gene chromatin in vitro. Cell 28:781–791.

Gray CP, Kaerner HC (1984): Sequence of the putative origin of replication in the U_L region of herpes simplex virus type 1 ANG DNA. J Gen Virol 65:2109–2119.

Guo Z-S, DePamphilis ML (1992): Specific transcription factors stimulate simian virus 40 and polyomavirus origins of DNA replication. Mol Cell Biol 12:2514–2524.

Hand R (1978): Eucaryotic DNA: Organization of the genome for replication. Cell 15:317–325.

Hatton KS, Dhar V, Brown EH, Iqbal MA, Stuart S, Didamo VT, Schildkraut CL (1988): Replication program of active and inactive multigene families in mammalian cells. Mol Cell Biol 8:2149–2158.

Hay RT, DePamphilis ML (1982): Initiation of SV40 DNA replication in vivo: Location and structure of 5′ ends of DNA synthesized in the ori region. Cell 28:767–779.

Held PG, Heintz NH (1992): Eukaryotic replication origins. Biochim Biophys Acta 1130:235–246.

Hendrickson EA, Fritze CE, Folk WR, DePamphilis ML (1987): The origin of bidirectional DNA replication in polyoma virus. EMBO J 6:2011–2018.

Hendrickson WG, Kusano T, Yamaki H, Balakrishnan R, King M, Murchie J, Schaechter M (1982): Binding of the origin of replication of *Escherichia coli* to the outer membrane. Cell 30:915–923.

Hobom G, Grosschedl R, Lusky M, Scherer G, Schwarz E, Kössel H (1979): Functional analysis of the replicator structure of lambdoid bacteriophage DNAs. Cold Spring Harb Symp Quant Biol 43:165–178.

Holmquist GP (1987): Role of replication time in the control of tissue-specific gene expression. Am J Hum Genet 40: 151–173.

Hsieh D-J, Camiolo SM, Yates JL (1993): Constitutive binding of EBNA1 protein to the Epstein-Barr virus replication origin, oriP, with distortion of DNA structure during latent infection. EMBO J 12:4933–4944.

Huberman JA (1991): Cell cycle control of initiation of eukaryotic DNA replication. Chromosoma 100:419–423.

Huberman JA, Riggs AD (1968): On the mechanism of DNA replication in mammalian chromosomes. J Mol Biol 32: 327–342.

Iguchi-Ariga SMM, Itani T, Kiji Y, Ariga H (1987a): Possible function of the *c-myc* product: Promotion of cellular DNA replication. EMBO J 6:2365–2371.

Iguchi-Ariga SMM, Itani T, Yamaguchi M, Ariga H (1987b): *c-myc* protein can be substituted for SV40 T antigen in SV40 DNA replication. Nucleic Acids Res 15:4889–4899.

Iguchi-Ariga SMM, Okazaki T, Itani T, Ogata M, Sato Y, Ariga H (1988): An initiation site of DNA replication with transcriptional enhancer activity present upstream of the *c-myc* gene. EMBO J 7:3135–3142.

Iguchi-Ariga SMM, Ogawa N, Ariga H (1993): Identification of the initiation region of DNA replication in the murine immunoglobulin heavy chain gene and possible function of the octamer motif as a putative DNA replication origin in mammalian cells. Biochim Biophys Acta 1172:73–81.

Johnston BH (1988): The S1-sensitive form of $d(C\text{-}T)_n \cdot d(A\text{-}G)_n$: Chemical evidence for a three-stranded structure in plasmids. Science 241:1800–1804.

Jones KA, Kadonaga JT, Rosenfeld PJ, Kelly TJ, Tjian R (1987): A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. Cell 48:79–89.

Jones KS, Tjian T (1984): Essential contact residues within SV40 large T antigen binding sites I and II identified by alkylation-interference. Cell 36:155–162.

Kimmerly W, Buchman A, Kornberg R, Rine J (1988): Roles of two DNA-binding factors in replication, segregation and transcriptional repression mediated by a yeast silencer. EMBO J 7:2241–2253.

Koepsel RR, Khan SA (1986): Statid and initiator protein-enhanced bending of DNA at a replication origin. Science 233:1316–1318.

Koff A, Schwedes JF, Tegtmeyer P (1991): Herpes simplex virus origin-binding protein (UL9) loops and distorts the viral replication origin. J Virol 65:3284–3292.

Kowalski D, Eddy MJ (1989): The DNA unwinding element: A novel, *cis*-acting component that facilitates opening of the *Escherichia coli* replication origin. EMBO J 8:4335–4344.

Kuo M-H, Grayhack E (1994): A library of yeast genomic MCM1 binding sites contains genes involved in cell cycle control, cell wall and membrane structure, and metabolism. Mol Cell Biol 14:348–359.

Kupershmidt S, Demarchi Lu Z, Ben-Porat T (1991): Analysis of an origin of DNA replication located at the L-terminus of the genome of pseudorabies virus. J Virol 65:6283–6291.

Landry S, Zannis-Hadjopoulos M (1991): Classes of autonomously replicating sequences are found among early-replicating monkey DNA. Biochim Biophys Acta 1088:234–244.

Leffak M, James CD (1989): Opposite replication polarity of the germ line c-*myc* gene in HeLa cells compared with that of two Burkitt lymphoma cell lines. Mol Cell Biol 9:586–593.

Leonhardt H, Page AW, Weier H-U, Bestor TH (1992): A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell 71:865–873.

Lin L-S, Meyer RJ (1987): DNA synthesis is initiated at two positions within the origin of replication of plasmid R1162. Nucleic Acids Res 15:8319–8331.

Linskens MHK, Huberman JA (1990): The two faces of higher eukaryotic DNA replication origins. Cell 62:845–847.

Løbner-Olesen A, Skarstad K, Hansen FG, von Meyenburg K, Boye E (1989): The DnaA protein determines the initiation mass of Escherichia coli K-12. Cell 57:881–889.

Martin ME, Piette J, Yaniv M, Tang W-J, Folk WR (1988): Activation of the polyomavirus enhancer by a murine activator protein 1 (AP1) homolog and two contiguous proteins. Proc Natl Acad Sci USA 85:5839–5843.

Mastrangelo IA, Hough PVC, Wall JS, Dodson M, Dean FB, Hurwitz J (1989): ATP-dependent assembly of double hexamers of SV40 T antigen at the viral origin of DNA replication. Nature 38:658–662.

Meijer M, Beck E, Hansen FG, Bergmans HEN, Messer W, von Meyenburg K, Schaller H (1979): Nucleotide sequence of the origin of replication of the *Escherichia coli* K-12 chromosome. Proc Natl Acad Sci USA 76:580–584.

Mukherjee S, Patel I, Bastia D (1985): Conformational changes in a replication origin induced by an initiator protein. Cell 43:189–197.

Mul YM, Verrijzer CP, van der Vliet PC (1990): Transcription factors NFI and NFIII/Oct-1 function independently, employing different mechanisms to enhance adenovirus DNA replication. J Virol 64:5510–5518.

Murakami Y, Satake M, Yamaguchi-Iwai Y, Sakai M, Muramatsu M, Ito Y (1991): The nuclear protooncogenes c-*jun* and c-*fos* as regulators of DNA replication. Proc Natl Acad Sci USA 88:3947–3951.

Newlon CS (1993): Two jobs for the origin repliation complex. Science 262:1830–1831.

Ogden GB, Pratt MJ, Schaechter M (1988): The replicative origin of the E. coli chromosome binds to cell membranes only when hemimethylated. Cell 54:127–135.

Oka A, Sugimoto K, Takanami M (1980): Replication origin of the *Escherichia coli* K-12 chromosome: The size and structure of the minimum DNA segment carrying the information for autonomous replication. Mol Gen Genet 178:9–20.

O'Neill EA, Fletcher C, Burrow CR, Heintz N, Roeder RG, Kelly TJ (1988): Transcription factor OTF-1 is functionally identical to the DNA replication factor NF-III. Science 241:1210–1213.

Palzkill TG, Newlon CS (1988): A yeast replication origin consists of multiple copies of a small conserved sequence. Cell 53:441–450.

Parsons R, Anderson ME, Tegtmeyer P (1990): Three domains in the simian virus 40 core origin orchestrate the binding, melting and DNA helicase activities of T antigen. J Virol 64:509–518.

Parsons RE, Stenger JE, Ray S, Welker R, Anderson ME, Tegtmeyer P (1991): Cooperative assembly of simian virus 40 T-antigen hexamers on fuctional halves of the replication origin. J Virol 65:2798–2806.

Passmore S, Maine GT, Elble R, Christ C, Tye B-K (1988): *Saccharomyces cerevisiae* protein involved in plasmid maintenance is necessary for mating of MATα cells. J Mol Biol 204:593–606.

Pearson CE, Ruiz MT, Price GB, Zannis-Hadjopoulos M (in press) Cruciform DNA-binding protein in HeLa cell extracts. Biochemistry 33:14185–14196.

Pruijn GJM, van Driel W, van der Vliet PC (1986): Nuclear factor III, a novel sequence-specific DNA-binding protein from HeLa cells stimulating adenovirus DNA replication. Nature 322:656–659.

Rao BS, Zannis-Hadjopoulos M, Price GB, Reitman M, Martin RG (1990): Sequence similarities among monkey ori-enriched (ors) fragments. Gene 87:233–242.

Reisman D, Yates J, Sugden B (1985): A putative origin of replication of plasmids derived from Epstein-Barr virus is composed of two *cis*-acting components. Mol Cell Biol 5:1822–1832.

Remington JA, Flickinger RA (1971): The time of DNA replication in relation to RNA synthesis in frog embryos. J Cell Physiol 77:411–422.

Russell DW, Zinder ND (1987): Hemimethylation prevents DNA replication in E. coli. Cell 1071–1079.

Saigo K, Yaginuma K, Hayashi H, Isono K, Takada S, and Koike K (1994) Enhanced expression of c-*myc* gene and binding protein to the 2.3 kb upstream region of the P1 promoter in human liver and colon cancers. Int J Oncol 5:1099–1104.

Santoro C, Mermod N, Andrews PC, Tjian R (1988): A family of human CCAAT-box-binding proteins active in transcription and DNA replication: Cloning and expression of multiple cDNAs. Nature 334:218–224.

Schnos M, Zahn K, Inman RB, Blattner FR (1988): Initiation protein induced helix destabilization at the λ origin: A prepriming step in DNA replication. Cell 52:385–395.

Schnos M, Zahn K, Blattner FR, Inman RB (1989): DNA looping induced by bacteriophage λ O protein: Implications for formation of higher order structures at the origin of replication. Virology 168:370–377.

Scott MP, Tamkun JW, Hartzell GW (1989): The structure and function of the homeodomain. Biochim Biophys Acta 989:25–48.

Seidman MM, Levine AJ, Weintraub H (1979): The asymmetric segregation of parental nucleosomes during chromosome replication. Cell 18:439–449.

Shinomiya T, Ina S (1993): DNA replication of histone gene repeats in *Drosophila melanogaster* tissue culture cells: Multiple initiation sites and replication pause sites. Mol Cell Biol 13:4098–4106.

Smithies O (1982): The control of globin and other eukaryotic genes. J Cell Physiol Suppl 1:137–143.

Snyder M, Buchman AR, Davis RW (1986): Bent DNA at a yeast autonomously replicating sequence. Nature 324:87–89.

Spradling A, Orr-Weaver T (1987): Regulation of DNA replication during *Drosophila* development. Annu Rev Genet 21:373–403.

Stow ND, McMonagle EC (1983): Characterization of the TR_s/IR_s origin of DNA replication of herpes simplex virus type 1. Virology 130:427–438.

Su W, Middleton T, Sugden B, Echols G (1991): DNA looping between the origin of replication of Epstein-Barr virus and its enhancer site: Stabilization of an origin complex with Epstein-Barr nuclear antigen 1. Proc Natl Acad Sci USA 88:10870–10874.

Subramanian KN, Shenk T (1977): Nucleotide sequence of a fragment of SV40 DNA that contains the origin of DNA replication and specifies the 5′ ends of "early" and "late" viral RNA. J Biol Chem 252:355–367.

Sugden B, Warren N (1989): A promoter of Epstein-Barr virus that can function during latent infection can be transactivated by EBNA-1, a viral protein required for viral DNA replication during latent infection. J Virol 63:2644–2649.

Taira T, Iguchi-Ariga, SMM, and Ariga H (1994) A novel DNA replication origin identified in the human heat shock protein 70 gene promoter. Mol Cell Biol 14:6386–6397.

Tasheva ES, Roufa DJ (1994a): A mammalian origin of bidirectional DNA replication within the Chinese hamster *RP14* locus. Mol Cell Biol 14:5628–5635.

Tasheva ES, Roufa DJ (1994b): Densely methylated DNA islands in mammalian chromosomal replication origins. Mol Cell Biol 14:5636–5644.

Taylor HJ (1984): Origins of replication and gene regulation. Mol Cell Biochem 61:99–109.

Tegtmeyer P, Lewton BA, DeLucia AL, Wilson VG, Ryder K (1983): Topography of simian virus 40 A protein-DNA complexes: Arrangement of protein bound to the origin of replication. J Virol 46:151–161.

Trempe JP, Lindstrom YI, Leffak M (1988): Opposite replication polarities of transcribed and nontranscribed histone H5 genes. Mol Cell Biol 8:1657–1663.

Tsurimoto T, Fairman MP, Stillman B (1989): Simian virus 40 DNA replication in vitro: Identification of multiple stages of initiation. Mol Cell Biol 9:3839–3849.

Umek RM, Kowalski D (1988): The ease of DNA unwinding as a determinant of initiation of yeast replication origins. Cell 52:559–567.

Vassilev L, Johnson EM (1990): An initiation zone of chromosomal DNA replication located upstream of the *c-myc* gene in proliferating HeLa Cells. Mol Cell Biol 10:4899–4904.

Vassilev LT, Burhans WC, DePamphilis ML (1990): Mapping an origin of DNA replication at a single-copy locus in exponentially proliferating mammalian cells. Mol Cell Biol 10:4685–4689.

Verrijzer CP, Kal AJ, van der Vliet PC (1990): The DNA binding domain (POU domain) of the transcription factor Oct-1 suffices for stimulation of DNA replication. EMBO J 9:1883–1888.

Villarreal LP (1991): Relationship of eukaryotic DNA replication to committed gene expression: General theory for gene control. Microbiol Rev 55:512–542.

Virshup DN (1990): DNA replication. Curr Opin Cell Biol 2:453–460.

Virta-Pearlman VJ, Gunaratne PH, Chinault AC (1993): Analysis of a replication initiation sequence from the adenosine deaminase region of the mouse genome. Mol Cell Biol 13:5931–5942.

Ward GK, McKenzie R, Zannis-Hadjopoulos M, Price GB (1990): The dynamic distribution and quantification of DNA cruciforms in eukaryotic nuclei. Exp Cell Res 188: 235–246.

Ward GK, Shihab-el-Deen A, Zannis-Hadjopoulos M, Price GB (1991): DNA cruciforms and the nuclear supporting structure. Exp Cell Res 195:92–98.

Wegner M, Zastrow G, Klavinius A, Schwender S, Muller F, Luksza H, Hoppe J, Wienberg J, Grummt F (1989): Cis-acting sequences from mouse rDNA promote plasmid DNA amplification and persistence in mouse cells: Implication of HMG-I in their function. Nucleic Acids Res 17: 9909–9932.

Weir HM, Stow ND (1990): Two binding sites for the herpes simplex virus type 1 UL9 protein are required for efficient activity of the oris replication origin. J Gen Virol 71:1379–1385.

Weir HM, Calder JM, Stow ND (1989): Binding of the herpes simplex virus type 1 UL9 gene product to an origin of DNA replication. Nucleic Acids Res 17:1409–1425.

Weller SK, Spadaro A, Schaffer JE, Murray AW, Maxam AM, Schaffer PA (1985): Cloning, sequencing, and functional analysis of ori_L, a herpes simplex virus type 1 origin of DNA synthesis. Mol Cell Biol 5:930–942.

Wu C, Zannis-Hadjopoulos M, Price GB (1993): In vivo activity for initiation of DNA replication resides in a transcribed region of the human genome. Biochim Biophys Acta 1174:258–266.

Wysokenski DA, Yates JL (1989): Multiple EBNA1-binding sites are required to form an EBNA1-dependent enhancer and to activate a minimal replicative origin with oriP of Epstein-Barr virus. J Virol 63:2657–2666.

Zannis-Hadjopoulos M, Kaufmann G, Wang S-S, Lechner RL, Karawya E, Hesse J, Martin RG (1985): Properties of some monkey DNA sequences obtained by a procedure that enriches for DNA replication origins. Mol Cell Biol 5:1621–1629.

Zannis-Hadjopoulos M, Frappier L, Khoury M, Price GB (1988): Effect of anti-cruciform DNA monoclonal antibodies on DNA replication. EMBO J 7:1837–1844.

Zhan K, Blattner FR (1985): Binding and bending of the λ replication origin by the phage O protein. EMBO J 4:3605–3616.

Zyskind JW, Smith DW (1986): The bacterial origin of replication, *oriC*. Cell 46:489–490.