A Revisionist Replicon Model for Higher Eukaryotic Genomes

J.L. Hamlin,* L.D. Mesner, O. Lar, R. Torres, S.V. Chodaparambil, and L. Wang

Department of Biochemistry and Molecular Genetics, University of Virginia School of Medicine, Charlottesville, Virginia

ABSTRACT

The replicon model devised to explain replication control in bacteria has served as the guiding paradigm in the search for origins of replication in the more complex genomes of eukaryotes. In *Saccharomyces cerevisiae*, this model has proved to be extremely useful, leading to the identification of specific genetic elements (replicators) and the interacting initiator proteins that activate them. However, replication control in organisms ranging from *Schizosaccharomyces pombe* to mammals is far more fluid: only a small number of origins seem to represent classic replicators, while the majority correspond to zones of inefficient, closely spaced start sites none of which are indispensable for origin activity. In addition, it is apparent that the epigenetic state of a given sequence largely determines its ability to be used as a replication initiation site. These conclusions were arrived at over a period of three decades, and required the development of several novel replicon mapping techniques, as well as new ways of examining the chromatin architecture of any sequence of interest. Recently, methods have been elaborated for isolating all of the active origins in the genomes of higher eukaryotes en masse. Microarray analyses and more recent high-throughput sequencing technology will allow all the origins to be mapped onto the chromosomes of any organism whose genome has been sequenced. With the advent of whole-genome studies on gene expression and chromatin composition, the field is now positioned to define both the genetic and epigenetic rules that govern origin activity. J. Cell. Biochem. 105: 321–329, 2008. © 2008 Wiley-Liss, Inc.

KEY WORDS: REPLICON; ORIGIN; REPLICATOR; CHROMATIN; TRANSCRIPTION

DNA fiber autoradiographic studies on yeast and mammalian genomes showed that origins of replication are bidirectional [Huberman and Riggs, 1968; Newlon et al., 1974] and are spaced, on average, 30 and 100 kb apart, respectively. Not surprisingly, the *replicon model* proposed in 1963 to explain the regulation of DNA replication in bacteria and their episomes [Jacob and Brenner, 1963] has been the guiding paradigm for virtually all subsequent studies directed toward the isolation and characterization of origins from eukaryotic genomes. In this model, a *replicon* was defined as a functional unit that contains two important elements: a gene encoding a *trans*-acting *initiator* and a mutable genetic *replicator* that directs the initiator (and initiation) to itself. Obviously, the model has had to be adjusted to accommodate the fact that eukaryotic chromosomes contain hundreds to thousands of individual origins [Huberman and Riggs, 1968; Newlon et al., 1974] that apparently are activated in each case by a single set of initiation factors.

The replicon model was so successful in guiding explorations in yeast that it promised to hold throughout the eukaryotic world. However, there has been a serious paradigm shift in the last decade or so because many aspects of the model do not adequately describe regulation of initiation in eukaryotic organisms more evolved than *Saccharomyces cerevisiae*. In this article, we will summarize the relevant data that have led to this paradigm shift. Our focus will be largely on the localization and characterization of DNA sequences that function as origins of replication in complex genomes—whether they be true *replicators* or not. There also has been an explosion of information in recent years concerning *initiators* in higher eukaryotic organisms, as well as epigenetic requirements for determining the activity of origins as well as the time of firing in the S-period. We refer the reader to excellent recent reviews that cover these important concepts [Bell and Dutta, 2002; Zhou et al., 2005; Aladjem et al., 2006], although we will briefly describe the state-of-the-art in these areas where relevant.

IDENTIFYING EUKARYOTIC REPLICATORS PHENOTYPICALLY

A critical advance in the eukaryotic replication field was the isolation of autonomously replicating sequence (ARS) elements from the *S. cerevisiae* and *Schizosaccharomyces pombe* genomes [Stinchcomb et al., 1979; Chan and Tye, 1980; Wright et al., 1986]. In *S. cerevisiae*, ARS elements are ∼100 bp in length and are characterized by a required consensus sequence (5′-[A/T]TTTAT[A/G]TTT[A/T]-3′) as well as two redundant, non-conserved, auxiliary elements [recently reviewed in Aladjem et al., 2006]. Identification of an ARS consensus allowed the isolation of the ORC complex [Bell and Stillman, 1992], which functions at origins as the initiator in this and apparently all other eukaryotes (although it has other non-origin targets). ARS elements from *S. pombe* are much larger (>500 bp) and consist of clusters of As and Ts that are asymmetrically distributed. Deletion of these elements in selected origins does, in fact, lower origin activity (reviewed in Kim and Huberman, 2001; Dai et al., 2005); however, there is no recognizable consensus element analogous to the *S. cerevisiae* ARS element, and unrelated AT-rich sequences can substitute in ARS assays without inactivating origin function. Thus, although ARS elements in *S. cerevisiae* and *S. pombe* both represent true, mutable, replicators, those in *S. pombe* clearly deviate from the original replicator paradigm.

Two-dimensional (2-D) gel replicon mapping methods that take advantage of the unique physical properties of restriction fragments containing either internal replication bubbles (i.e., origins), single forks, or termination structures [Brewer and Fangman, 1987; Nawotka and Huberman, 1988] have shown that most *S. cerevisiae* ARS elements correspond to legitimate chromosomal origins [reviewed in Aladjem et al., 2006]. In both yeasts, most origins are not active 100% of the time, with some being activated in fewer than 10% of cell cycles. Thus, *bona fide* termini defining individual replicons are not likely to be common in these genomes. A corollary to this inefficiency is that closely spaced potential initiation sites are unlikely to interfere with each other by firing on the same template, as has been reported for very active origins juxtaposed on plasmids [Brewer and Fangman, 1993].

The ARS assay, when applied to complex mammalian genomes, has not led to the identification of verifiable genetic replicators. For example, when a cloned human genomic library was transfected into a human cell line [Krysan et al., 1989], virtually every cloned fragment replicated autonomously to some degree (even those of bacterial origin), but the larger fragments were much more efficient. This result suggested either that origins of replication are distributed at very frequent intervals in mammalian genomes, or that virtually any sequence, when taken out of context, can serve as a template for initiation at a measurable frequency. These suggestions were compatible with observations that initiation occurs at closely spaced, apparently random, sequences in the early cleavage stages of *Xenopus laevis* and *Drosophila melanogaster* development [e.g., Blumenthal et al., 1974]. Furthermore, virtually any DNA sequence (even from viruses and bacterial plasmids) can replicate in vitro in frog egg extracts prior to the mid-blastula transition when transcription begins [Harland and Laskey, 1980].

IDENTIFYING ORIGINS AND POTENTIAL REPLICATORS BY BIOCHEMICAL METHODS THAT LOCALIZE START SITES

Given the lack of a useful ARS assay for rescuing higher eukaryotic replicators, identification of origins has relied on the biochemical approach. The several origin mapping strategies all have required that the region under study be cloned and mapped. As a consequence, the majority of chromosomal regions were characterized initially because of previous interest in their respective genes. Methods for origin localization fall into several categories [reviewed in Hamlin, 1992]: (1) labeling nascent strands in early S-phase with radioactive thymidine or BUdR, which are then used to detect the earliest fragments synthesized in a region of interest; (2) monitoring the direction of fork movement through a defined region; (3) determining the position at which either leading or lagging strands switch from one template to the other; or (4) monitoring the development and expansion of replication bubbles around a nascent strand start site (Fig. 1).

The first mammalian origin was identified in the 240 kb amplified dihydrofolate reductase (DHFR) domain of the methotrexate-resistant CHO cell line, CHOC 400. This locus has been analyzed by early-labeled fragment (ELF) assays, by leading and lagging strand template bias assays, by PCR-based small nascent strand abundance assays, and by both the neutral/neutral and neutral/alkaline 2-D gel assays [reviewed in Dijkwel and Hamlin, 1996]. Virtually all of the studies on this complex origin have converged on a model in which initiation can occur at more than 30 sites within the 55 kb spacer between the convergently transcribed DHFR and 2BE2121 genes, with the efficiencies of utilization varying from a minimum near the 3′ ends of the convergently transcribed genes to maxima at two different sites (termed *ori-β* and *ori-γ*), which are spaced about 20 kb apart [e.g., Leu and Hamlin, 1989]. A quantitative estimate suggests that, at most, 20% of initiations in the 55-kb spacer occur within the 2 kb encompassing the most active *ori-β* region, and the spacer as a whole sustains an initiation event in only 15–20% of cell cycles. Thus, even the most active region in this origin (*ori-β*) is very inefficient, firing in only 3–4% of cell cycles.

How do these results on the prototypic DHFR origin compare to other origins that have been identified by locus-centric approaches in higher eukaryotes? By one (usually) or more (rarely) mapping methods, about 30 origins have now been identified in organisms ranging from fruit flies to man [see Aladjem et al., 2006 for original references]. Origins have been localized in the amplified *D. melanogaster* chorion locus, *Xenopus laevis* and human rDNA loci, and the Syrian hamster CAD and murine adenosine deaminase domains. *Single-copy* human origins have been identified in the c-myc, β-globin, IGFII, HSP70, MCM4/PRKDC, RPE, TOP1, and YWHAH domains. Other single-copy origins have been localized in *Drosophila* histone and polymerase-β domains, in African green monkey cells, in the murine β-globin and IGH loci, as well as in the Chinese hamster rhodopsin, APRT, GNA13, GADD45A, RSP14, and TK1 domains.

Many of these origins have been shown on 2-D gels, or by extensive analysis of the surrounding region by the small nascent
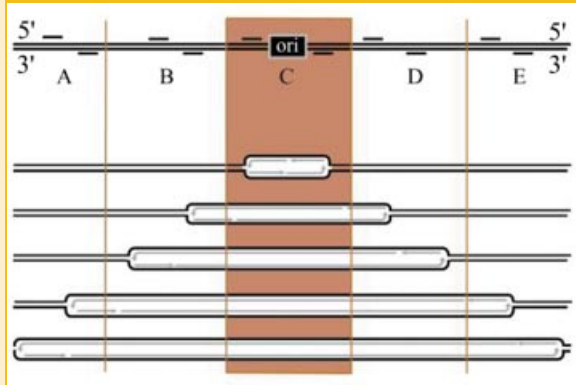
Fig. 1. Methods of localizing origins of replication. The cartoon depicts five adjacent restriction fragments (A–E) with the central one containing an origin. The inner gray lines represent the leading and lagging nascent strands as the bubble matures. The methods for identifying fragment C fall in to four general categories. 1) Determining the positions of the small nascent strands contained within the bubbles (inner gray lines); this can be done by labeling them in vivo with bromodeoxyuridine (BrdU) or radioactive thymidine shortly after initiation and somehow identifying their origins; alternatively small nascent strands are melted away from template and isolated by size; in the most common recent application, the small melted nascent strands are quantified by PCR, using a set of primer pairs positioned throughout the region of interest (shown as pairs of lines flanking the linear map above); this method works well on single-copy sequences owing to the great sensitivity of PCR (see Giacca et al., 1994). 2) Relying on the unique electrophoretic mobilities of restriction fragments containing either replication bubbles, single forks, or X-shaped termination structures under appropriate conditions (2-D replicon methods); genomic DNA is gently isolated, digested with an appropriate restriction enzyme, and separated on a neutral/neutral 2-D gel; after transfer of the digest to a membrane, the intermediates are hybridized successively with a series of probes from the region of interest (see Brewer and Fangman, 1987). 3) Determining the direction of replication fork movement through a region of interest, which should diverge from an origin; the neutral/ alkaline 2-D gel method (Nawotka and Huberman, 1988) and the more recent SMARD technique that examines DNA fibers from cells pulsed with two different nucleotides, which are then detected with fluorescent probes of different colors (e.g., Norio et al., 2005). 4) Determining the position at which leading or lagging strands switch template strands, which occurs at origins of replication (Handeli et al., 1989; Burhans et al., 1990; Dijkwel et al., 2002). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

strand abundance assay, to represent zones of inefficient initiation sites (although certain subregions can be preferred). Importantly, most of the origins that have been identified in the last 8–10 years were discovered with the PCR-based small nascent strand abundance assay [e.g., Giacca et al., 1994], and most of these have not been characterized with any other state-of-the-art assay such as neutral/neutral or neutral/alkaline 2-D gel analysis, either of which can readily distinguish isolated single initiation sites from sites residing within initiation zones. Indeed, our laboratory has shown that the small nascent strand abundance assay is insensitive to low levels of initiation that can be detected on 2-D gels [Dijkwel et al., 2002]. Thus, the jury is still out as to how many of the 30 higher eukaryotic origins actually reside in initiation zones.

Apparently, the vast majority of active origins in *S. cerevisiae* [Raghuraman et al., 2001] and in *S. pombe* [Gomez and Antequera, 1999] reside in the spacers between genes. All but five of the

mammalian origins cited above (Syrian hamster CAD; Chinese hamster GADD45A, and RSP14, as well as human DMMT1 and YWHAH) reside in intergenic regions [see Aladjem et al., 2006 for original references]. This is consistent with the observation that in early cleavage stages in *Xenopus*, initiation can occur in both the transcription units and spacers when the genes are silent prior to the mid-blastula transition, but only in the spacers after transcription commences [Hyrien et al., 1995]. In concert with older electron microscopic studies on early cleavage *Xenopus* and *Drosophila* embryos [e.g., Blumenthal et al., 1974; Buongiorno-Nardelli et al., 1976], these data suggest that, at least in early development in these organisms, the entire genome constitutes a potential substrate for initiation.

However, other higher eukaryotic origins are clearly more circumscribed and come closer to the bacterial paradigms that generated the replicon model in the first place. These include the human lamin B2 [Giacca et al., 1994] and β-like globin [Aladjem et al., 1998] origins, possibly the Chinese hamster GNA13 origin [Toledo et al., 1998], and the *ardB&ardC* [Benard et al., 1996] and rDNA origins in *Physarum* [Benard et al., 1995]. Of these, the human globin and GNA13 origins, as well as the *Physarum* rDNA origin, constitute mini-zones of initiation, while the lamin B2 [Abdura-shidova et al., 2000] and *ardB&C* [Benard et al., 1996] origins appear to correspond most closely to single sites. Interestingly, the latter two reside within the promoters of the nearby genes. With the exception of the *Physarum* origins and Chinese hamster GNA13, however, none of these has subjected to 2-D gel analysis.

Thus, there appears to be a spectrum of origin types ranging from quite circumscribed to extremely broad initiation zones. Likewise, the efficiency of origin firing apparently can range from nearly 100% (e.g., human β-globin [Kitsberg et al., 1993] and *Physarum* ardB&C [Benard et al., 1996]) to less than 5% [e.g., the CHO DHFR; *ori-β* locus; Dijkwel et al., 2002].

Recently, several groups have devised methods for isolating origins from complex genomes *en masse*, in an attempt to discover how analyses of the 30 or so origins cited (most of which were localized by virtue of proximity to one's favorite gene) reflect the spectrum of origins in the genome as a whole. The first such study isolated nascent strands 800–3,000 nt in length and prepared a small library that was shown to be enriched ~7-fold over the starting total genomic DNA [Todorovic et al., 2005]. Although several new origins were identified in this study, there apparently was a high degree of contamination with non-origin material. A second approach takes advantage of the circular nature of restriction fragments containing internal initiation sites (i.e., replication bubbles or eye forms) by trapping them in gelling agarose [Mesner et al., 2006]. Contaminat-ing simple Y structures, linear fragments, and X-shaped termination structures then can be electrophoresed out of the agarose plug, leaving purified bubble-containing fragments behind. In a pilot project, this method yielded thousands of clones from CHO cells, virtually all of which appear to correspond to initiation sites in the hamster genome [Mesner et al., 2006], as demonstrated by 2-D gel analysis of their respective genomic positions. Importantly, >90% of the cloned genomic regions that were analyzed in early S-phase cells display the composite pattern characteristic of initiation sites residing in initiation zones [i.e., a complete bubble arc and a

complete, more prominent single fork arc; L.D. Mesner, unpublished work; Mesner et al., 2006]. However, the bubble-to-fork-arc ratios varied greatly among them, suggesting that they arise from both large and small initiation zones. The 2-D gel analysis also suggested that most if not all origins probably fire in less than 30% of cell cycles.

In a more recent study, libraries of millions of clones from human cells have been prepared and used to probe microarrays of the 30 MB of genomic DNA being analyzed in the NHGRI ENCODE project (N. Karnani, L.D. Mesner, C. Taylor, A. Malhotra, J. L. Hamlin, and A. Dutta, unpublished work). This method, too, suggests that more than 50% of the cloned bubble-containing fragments arise from initiation zones. This number is likely to be even larger, because many of the smaller zones will be confined to a single restriction fragment, as would be the case with the human c-myc and β-globin origins. Thus, our current thinking is that most mammalian origins correspond to initiation zones rather than single sites analogous to the lamin B2 origin. By hybridizing these libraries to microarrays and comparing their distributions to the other information being compiled for the 30 MB ENCODE region, it soon should be possible to paint a relatively detailed picture of the activity of origins *vis-a-vis* their times of synthesis, proximity to active genes, chromosomal *milieu* (i.e., histone and DNA covalent modifications), etc.

IN SEARCH OF CLASSIC GENETIC REPLICATORS IN HIGHER EUKARYOTIC GENOMES

Huberman and Riggs [1968] coined the operational term *origin* to denote replication initiation sites in complex genomes. The term *replicator* arose directly from the replicon model of Jacob and Brenner [1963], and is defined as a mutable genetic controlling element. Origins necessarily coincide with replicators in simple genomes with fixed initiation sites (bacteria, viruses and plasmids, budding yeast), and it has been assumed that the multiple origins in complex genomes would do the same. However, as pointed out above, the ARS assay has not been successful in identifying replicators from complex genomes more evolved than *S. pombe.* Therefore, small fragments containing origins identified by biochemical means have been analyzed for the presence of required genetic elements in two other ways: (1) by positioning them (or mutated versions) at random or fixed ectopic chromosomal sites to ask whether initiation competence travels with the fragment (the negative control is usually a nearby site in the accompanying vector), and (2) by performing *in loco* mutagenesis by homologous recombination to determine whether initiation sites are required genetic elements.

By the first approach, mutable "replicator" activity has been detected in the ACE origin from the *Drosophila* chorion puff [Delidakis and Kafatos, 1987; Heck and Spradling, 1990], the human β-globin origin [Aladjem et al., 1998], the human c-myc origin [Liu et al., 2003], the human lamin B2 origin [Paixao et al., 2004], and the CHO *ori-β* locus [e.g., Handeli et al., 1989; Gray et al., 2007]. Of these, both the human β-globin origin [Aladjem et al., 1998] and the human c-myc origin [Liu et al., 2003] appear to have redundant sequences that must both be removed to inhibit origin activity;

additionally, the β-globin, c-myc, *ori-β*, and ACE sequences all derive from zones of initiation sites in their native chromosomal contexts. Note, however, that initiation does not occur within ACE itself. Thus, only the very AT-rich lamin B2 origin, which is in a very narrow intergenic spacer, appears to correspond to a single, mutable origin/replicator by this criterion. Furthermore, a subregion of this fragment has been shown in ChIP assays to bind to hORC2 [reviewed in Aladjem et al., 2006]. Interestingly, the same techniques used to identify a relatively fixed origin containing mutable genetic elements in the human β-like globin locus detected no such circumscribed start sites or potential replicators in the homologous region in the murine genome [see Aladjem et al., 2006]. Thus, the mechanisms of initiation at these loci seem not to have been conserved between the two species.

The DHFR initiation zone constitutes one of the few origins that has been mutagenized systematically *in loco* to search for required genetic elements (replicators). Although the activity of the *ori-β* locus can be compromised by mutagenesis when placed by itself at ectopic sites [Gray et al., 2007], deletion from its *in loco* position has no effect on initiation in the remainder of the 55 kb initiation zone nor on the time of replication of the locus as a whole [Kalejta et al., 1998]. In fact, the central 45 kb of the intergenic spacer, which encompasses >95% of the usual nascent strand start sites in this zone, can be deleted without a significant change in the early-replicating properties of the DHFR locus [Mesner et al., 2003b]; this was achieved by greatly increasing the frequency of initiation in the truncated spacer that remained. However, it remains to perform the same systematic genetic analysis that was carried out at ectopic positions on the *ori-β* region *in loco*. If the same results were obtained, one could conclude that, *ori-β* is, indeed, a genetic replicator, but it must be redundant with many other replicators spread throughout the intergenic spacer. Alternatively, a responsible replicator could conceivably reside outside of the actual initiation zone itself.

Interestingly, when tested directly, a fragment containing *ori-β* proved to replicate autonomously in human or CHO cells no better than similarly-sized fragments that presumably do not contain replicators [P. Foreman, J.D. Milbrandt, and J.L. Hamlin, unpublished work; Caddle and Calos, 1992]. (Parenthetically, the notion that matrix attachment regions (MARs) might coincide with genetic replicators [Razin et al., 1986] was tested directly by deleting the prominent MAR that lies in the approximate center of the DHFR initiation zone [Mesner et al., 2003a]. However, its removal had no effect on initiation per se; rather, the daughter chromatids, which are usually separated from one another locally within minutes of replication, failed to separate until just prior to mitosis. Thus, at least in this locus, the local MAR is within the origin but is not the element responsible for origin activity.)

EPIGENETIC FACTORS THAT REGULATE ORIGIN ACTIVITY AND THE TIME OF FIRING

Is it possible that virtually any sequence in complex genomes can serve as a template for initiation provided that it is in a permissive environment? There are several examples of the

plasticity of origins of replication that support this view. For example, when nucleotide pools are lowered in CHO cells by inhibitors [Anglana et al., 2003], or when the transcription patterns of neighboring genes are altered during development, the number and distribution of initiation sites can change dramatically [e.g., Norio et al., 2005]. These findings are compatible with the idea that there are many more potential initiation sites in the genome than are normally used in a given cell cycle, but their usage can be negatively or positively regulated by environmental effects such as local gene activity and/or chromatin architecture.

In support of these ideas are the following examples. Transcription through *S. cerevisiae* ARS1 [Snyder et al., 1988], as well as through cloned human fragments [Haase et al., 1994], inhibits their ability to replicate autonomously. In the CHO DHFR locus, deletion of the 3′ processing signals from the DHFR gene allows RNA polymerase to traverse almost the entire former intergenic spacer, completely eliminating replication initiation in its path [Mesner and Hamlin, 2005]. Conversely, when a fragment from the body of the DHFR gene, which never initiates *in loco*, is inserted at ectopic (presumably non-transcribed) chromosomal sites, it now serves as a perfectly good template for replication initiation [Lin et al., 2005]. Interestingly, deletion of a functional promoter from the DHFR gene, which prevents transcription by RNA polymerase, allows the former 55 kb initiation zone to spread into the body of the now inactive gene [Saha et al., 2004]. These data again suggest that potential initiation sites are spread throughout the genome — even in the bodies of potential genes, but their use is inhibited by transcription through the template. As noted above, however, at least five mammalian origins reside in the coding regions of active genes. It remains to be determined whether both alleles at these loci are actively transcribing. Indeed, whole-genome microarray approaches that identified initiation sites in *S. cerevisiae, S. pombe*, and *D. melanogaster*, suggest that most sites reside in non-transcribed spacers [e.g., Raghuraman et al., 2001; MacAlpine et al., 2004].

In contrast to the negative effects of transcription through a DNA sequence on replication initiation, several recent studies suggest that proximity to an active promoter can establish a chromatin ambience that facilitates origin activity without requiring transcription itself. In *S. cerevisiae*, both RNA Polymerase II and III transcription factors can stimulate initiation at the nearby ARS1, apparently by facilitating chromatin remodeling [Bodmer-Glavas et al., 2001]. In *Xenopus laevis* eggs, an artificial origin can be created in a plasmid simply by targeting Gal4-VP16 to binding sites positioned next to a TATA box [Danis et al., 2004]. Transcription is not required for this effect, and origin activation is accompanied by local histone acetylation. Recent studies on the c-myc replicator suggest that local promoter elements are required for origin activity, but that very active transcription from that promoter inhibits origin firing [Ghosh et al., 2004].

Additional studies on *D. melanogaster* ACE (the amplification control element) show that origins of replication in this system bind well-known transcription factors, suggesting a direct role for these factors in initiation of replication [Bosco et al., 2001; Beall et al., 2002]. The lamin B2 origin resides in a very narrow intergenic zone, and elements that enhance transcription (e.g., a CpG island) also enhance origin activity [Paixao et al., 2004]. In fact, sequences encompassing CpG islands are greatly enriched in small nascent strands in mammalian cells [Delgado et al., 1998], again suggesting that transcription factors can modulate both gene and local origin activity. Various attempts have been made to categorize the elements residing in or near origins of replication in mammalian cells, and although there is no consistent pattern among the various well-studied examples, transcription factor binding sites appear regularly in or near several origins [Aladjem et al., 2006]. In most cases, however, the precision with which replication initiation sites have been mapped does not allow the conclusion that these sites have anything to do with origin activity per se.

Another clear-cut example of promoter/origin interactions is manifested in the CHO DHFR locus. The DHFR promoter lies 26 kb upstream from the nearest initiation site in the intergenic spacer, yet deletion of critical elements that eliminate transcription greatly diminishes the efficiency of initiation of replication in the spacer [Saha et al., 2004]. It has not yet been determined whether inhibition of origin activity by the promoter deletion results from the absence of transcription per se, as opposed to some transcription-independent change in chromatin architecture that depends on an intact promoter and extends all the way into the intergenic zone. This effect may be manifested on a more global scale by studies in which nuclei from CHO cells at various positions in the cell cycle are allowed to replicate in *Xenopus* egg extracts. These nuclei have been shown to pass through a regulatory origin decision point (ODP) in early G_1 that converts a totally random pattern of initiation throughout the DHFR locus (and presumably the whole genome) to the more focused intergenic pattern displayed by cultured cells [e.g., Keezer and Gilbert, 2002]. There is some evidence to suggest that the critical difference between early- and late-G_1 nuclei is the dramatic increase in transcription of neighboring genes that occurs during this interval. However, transcription commences in this system early in G_1, well before the ODP itself. Clearly, something else has to happen in addition to transcription per se to activate the nearby origin.

Thus, there appear to be both local and more far-reaching effects of transcription competence on origin activity, neither of which may require transcription itself under most circumstances. Data from a genome-wide microarray study on replication timing, transcription patterns, and the distribution of ORC in the *Drosophila* genome may suggest how this is achieved [MacAlpine et al., 2004]. Based on the interaction of these different, large, data sets, it appears that transcription (or a competent promoter) may have a local effect by influencing origin selection, while at the same time regulating origin activation over domains at least 100 kb in length (in the latter case by determining the activation of or the time of activation of origins within the domain). In fact, a single molecule analysis of the multicopy rDNA locus in *S. cerevisiae* reveals that the intergenic rDNA origins fire in clusters [Pasero et al., 2002], which could result from packaging several tandem copies of the repeating unit into distinct architectural chromatin domains. Presumably, in all of these cases, transcription factors are mediating chromatin remodeling toward a permissive configuration. The question then arises as to what constitutes a permissive chromatin ambience. There is little doubt that most active origins will find themselves in euchromatin,

which has long been known to replicate early in S-phase [Goldman et al., 1984]. Recent genome-wide studies on covalent histone modifications in *Drosophila* define a binary pattern in which active genes are hyperacetylated on histones H3 and H4 and hypermethylated at H3/K4 and H3/K79; inactive genes, on the other hand, are hypomethylated and deacetylated at the same lysines [Schubeler et al., 2004]. These modifications are tightly linked to RNA polymerase activity itself. However, this suggestion is not consistent with studies cited above in which the act of transcription (rather than transcription factor loading at the promoter) is not required to activate the local origins. Furthermore, most activating marks are prevalent only in the promoter and immediately downstream [Pattenden et al., 2005].This suggests that acetylation of histones in promoters could be responsible for activating those origins that lie within a short distance (e.g., c-myc, β-globin, lamin B2), but cannot readily explain activation of origins that lie considerable distances from a local promoter (e.g., the DHFR origin). Therefore, no unifying rules have been uncovered to explain the complex and subtle effects of transcription (or active promoters) on establishing a permissive environment for initiation.

Altogether, these data are consistent with earlier studies showing that actively transcribed genes in somatic metazoan cells are very often replicated in early S-phase when most of the origins are likely to be activated [e.g., Holmquist, 1987]. In budding yeast, the default time of activation of most origins generally appears to be early S-phase, although the activity and/or time of firing of ARS elements can be regulated by chromosomal context effects such as proximity to telomeres [Ferguson and Fangman, 1992].

Indeed, there is a well-established literature suggesting that, at a population level, origins of replication in budding yeast [e.g., Raghuraman et al., 2001], fission yeast [e.g., Segurado et al., 2003], fruit flies [e.g., MacAlpine and Bell, 2005], and mammals [e.g., Jeon et al., 2005] fire in predetermined orders within the S-phase. However, in a recent analysis of replication timing of single DNA fibers in *S. cerevisiae*, it is quite clear that the timing of replication of given chromosomal segments is essentially probabilistic within an individual cell and is not strictly determined [Czajkowsky et al., 2008]; however, the averaged replication times for all molecules examined recapitulates the ensemble-averaged patterns observed in the population. This same phenomenon may be reflected in the fact that the single-copy, so-called early-firing, mammalian origins that have been examined on 2-D gels clearly fire at any time during a 2–3 h window—presumably also in a probabilistic manner [Dijkwel and Hamlin, 1992]. In the few documented cases of mammalian mid- or late-firing origins that have been examined on 2-D gels, the window of activation is equally broad [e.g., Larner et al., 1999]. Even in synchronized populations of budding yeast, the activation times of early- and mid-firing origins, as judged from 2-D gels, clearly overlap [Ferguson and Fangman, 1992], suggesting the absence of any strictly defined temporal organization.

With these caveats in mind, a well-studied example of regulation of the time of firing is the β-globin domain in humans, in which there is a clearly defined upstream locus control region (LCR) whose loss in the Hispanic deletion renders the downstream origin late-firing [Aladjem et al., 1995]. Another important contribution shows

that tethering a histone acetyl transferase to a late-firing origin in *S. cerevisiae* can advance its time of firing in the S-period [Vogelauer et al., 2002]. Thus, the transcription apparatus may serve only as a conveyor of the enzymes and factors required to adequately remodel or modify chromatin in the vicinity of a potential origin. Interestingly, there is another way of establishing replication timing, which was uncovered in a study of the EBV origin of replication. This origin utilizes the viral EBNA protein in addition to cellular proteins to effect initiation in a once-per-cell-cycle mode of replication [Zhou et al., 2006]. The nucleosomes that flank the dyad symmetry (DS) element in this origin are remodeled by SNF2H in late G_1 prior to S-phase entry. At the same time, and unlike other examples of origin activation, histones in the EBV origin are *deacetylated* by HDAC2. The authors point out that this origin is probably late-firing, so deacetylation may be the mark that delays origin firing in this system.

Regardless of how replication timing is established, a fascinating recent study suggests that *transcriptional* competence is established at the time of replication (i.e., exogenous genes are better able to support transcription when injected into early S-phase nuclei than when delivered in late S-phase [Zhang et al., 2002]). These studies suggest how epigenetic states can be maintained from one cell generation to another. For a comprehensive summary of this fascinating aspect of origin regulation, the reader is directed to a recent review of replication timing related to imprinting and X-chromosome inactivation [Lande-Diner and Cedar, 2005]. Lastly, another potential epigenetic influence that deserves further study is methylation of CpG islands. The jury is still out on the real involvement of this covalent DNA modification in origin activation.

How do these data fit the original replicon model, and does that model help or hinder our future ability to understand regulation of DNA synthesis in higher eukaryotes? We would like to propose the following model for control of replication in higher eukaryotic cells, which is fluid enough to accommodate the overwhelming majority of the complex data sets cited above. We suggest that the genomes of metazoans are peppered at very frequent but random intervals (perhaps every few hundred base pairs) with a hierarchy of potential, degenerate, replicators. When activated, these replicators are suggested to control initiation only in their immediate environments. In the theoretical absence of any chromosomal context effects (i.e., in naked DNA), these sites could be more or less active as recognition elements for metazoan initiators in much the same way that micrococcal nuclease exhibits a hierarchic preference for certain sequences in naked DNA when enzyme is limiting. Clearly, when viewed in isolation, deletion or mutagenesis of such an initiation (or cleavage) site would lower or eliminate its ability to attract the initiation complex (or nuclease) and inhibit initiation (or cleavage) at that site, but not at neighboring sites. In general, most of these initiation sites are suggested to be highly inefficient, explaining why even broad initiation zones appear to sustain only one or fewer initiation events within the zone in a given cell cycle.

Thus, the probability that any given site will efficiently attract an initiation complex will depend upon whether it finds itself in a permissive environment. In a euchromatic chromatin domain, the overall architecture would be permissive, allowing a non-transcribed (intergenic) region to initiate with more or less efficiency

depending on its complement of other proteins (e.g., transcription factors, modified histones, etc.), while sites within an active transcription unit in the same euchromatic domain would not. The non-permissive environment of heterochromatin would preclude initiation of replication, and such regions would have to wait to be replicated passively from active upstream or downstream euchromatic regions, probably by a preceding wave of de-heterochromatization.

In this general model, lengthy intergenic spacers such as DHFR and rhodopsin correspond to broad initiation zones because they contain many exposed degenerate replicators, some of them more efficient than others. Conversely, very narrow intergenic spacers might isolate only one or a few active initiation sites (as in the lamin B2 and human β-globin loci), even though other inchoate sites might reside in the bodies of the neighboring actively-transcribed genes that define the spacer. Any of these sites (intra- or intergenic) might behave as classic, mutable, replicators when positioned at a permissive ectopic site. This general model could be extended to suggest that the one or two solo sites in narrow spacers may have evolved particularly high affinities for the initiation complex in order to ensure that they are activated in every cell cycle.

Viewed in this way, one could question the value of mutagenizing a site such as lamin B2 or ori-β and testing the effects on initiation only within the immediate environment of a cassette at an ectopic chromosomal site [e.g., Aladjem et al., 1998; Altman and Fanning, 2004]. It is already clear that there are no recognizable consensus sequences among the known origins except for the usual presence of AT-rich and easily unwound elements in the neighborhood [reviewed in Gilbert, 2004; Aladjem et al., 2006]. However, these mutagenic studies will likely define the nature of sequence *characteristics* that have increased affinity for ORC and/or other proteins involved in the initial melting of the helix. To be truly meaningful, however, these approaches will have to control for possible effects of mutations on other epigenetic phenomena that play important roles in origin activity (e.g., local transcription or chromatin modification).

Perhaps by modulating our thinking around this more inclusive general model, our experiments can be designed to give the maximum amount of information about this elemental but complex process.

ACKNOWLEDGMENTS

REFERENCES

Abdurashidova G, Deganuto M, Klima R, Riva S, Biamonti G, Giacca M, Falaschi A. 2000. Start sites of bidirectional DNA synthesis at the human lamin B2 origin. Science 287:2023–2026.

Aladjem MI, Groudine M, Brody LL, Dieken ES, Fournier RE, Wahl GM, Epner EM. 1995. Participation of the human beta-globin locus control region in initiation of DNA replication. Science 270:815–819.

Aladjem MI, Rodewald LW, Kolman JL, Wahl GM. 1998. Genetic dissection of a mammalian replicator in the human beta-globin locus. Science 281:1005–1009.

Aladjem MI, Falaschi A, Kowalski D. 2006. Eukaryotic DNA replication origins. DePamphilis M, editor. DNA replication and human disease. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. pp 31–62.

Altman AL, Fanning E. 2004. Defined sequence modules and an architectural element cooperate to promote initiation at an ectopic mammalian chromosomal replication origin. Mol Cell Biol 24:4138–4150.

Anglana M, Apiou F, Bensimon A, Debatisse M. 2003. Dynamics of DNA replication in mammalian somatic cells: Nucleotide pool modulates origin choice and interorigin spacing. Cell 114:385–394.

Beall EL, Manak JR, Zhou S, Bell M, Lipsick JS, Botchan MR. 2002. Role for a Drosophila Myb-containing protein complex in site-specific DNA replication. Nature 420:833–837.

Bell SP, Dutta A. 2002. DNA replication in eukaryotic cells. Annu Rev Biochem 71:333–374.

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

Benard M, Lagnel C, Pierron G. 1995. Site-specific initiation of DNA replication within the non-transcribed spacer of Physarum rDNA. Nucl Acids Res 23:1447–1453.

Benard M, Lagnel C, Pallotta D, Pierron G. 1996. Mapping of a replication origin within the promoter region of two unlinked, abundantly transcribed actin genes of Physarum polycephalum. Mol Cell Biol 16:968–976.

Blumenthal AB, Kriegstein HJ, Hogness DS. 1974. The units of DNA replication in Drosophila melanogaster chromosomes. Cold Spring Harb Symp Quant Biol 38:205–223.

Bodmer-Glavas M, Edler K, Barberis A. 2001. RNA polymerase II and III transcription factors can stimulate DNA replication by modifying origin chromatin structures. Nucl Acids Res 29:4570–4580.

Bosco G, Du W, Orr-Weaver TL. 2001. DNA replication control through interaction of E2F-RB and the origin recognition complex. Nat Cell Biol 3:289–295.

Brewer BJ, Fangman WL. 1987. The localization of replication origins on ARS plasmids in S cerevisiae. Cell 51:463–471.

Brewer BJ, Fangman WL. 1993. Initiation at closely spaced replication origins in a yeast chromosome. Science 262:1728–1731.

Buongiorno-Nardelli M, Amaldi F, Lava-Sanchez PA. 1976. Electron microscope analysis of amplifying ribosomal DNA from Xenopus laevis. Exp Cell Res 98:95–103.

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

Caddle MS, Calos MP. 1992. Analysis of the autonomous replication behavior in human cells of the dihydrofolate reductase putative chromosomal origin of replication. Nucl Acids Res 20:5971–5978.

Chan CS, Tye BK. 1980. Autonomously replicating sequences in Saccharomyces cerevisiae. Proc Natl Acad Sci USA 77:6329–6333.

Czajkowsky DM, Liu J, Hamlin JL, Shao Z. 2008. DNA combing reveals intrinsic temporal disorder in the replication of yeast chromosome VI. J Mol Biol 375:12–19.

Dai J, Chuang RY, Kelly TJ. 2005. DNA replication origins in the Schizosaccharomyces pombe genome. Proc Natl Acad Sci USA 102:337–342.

Danis E, Brodolin K, Menut S, Maiorano D, Girard-Reydet C, Mechali M. 2004. Specification of a DNA replication origin by a transcription complex. Nat Cell Biol 6:721–730.

Delgado S, Gomez M, Bird A, Antequera F. 1998. Initiation of DNA replication at CpG islands in mammalian chromosomes. EMBO J 17:2426–2435.

Delidakis C, Kafatos FC. 1987. Amplification of a chorion gene cluster in Drosophila is subject to multiple cis-regulatory elements and to long-range position effects. J Mol Biol 197:11–26.

Dijkwel PA, Hamlin JL. 1992. Initiation of DNA replication in the dihydrofolate reductase locus is confined to the early S period in CHO cells synchronized with the plant amino acid mimosine. Mol Cell Biol 12:3715–3722.

Dijkwel PA, Hamlin JL. 1996. Sequence and context effects on origin function in mammalian cells. J Cell Biochem 62:210–222.

Dijkwel PA, Wang S, Hamlin JL. 2002. Initiation sites are distributed at frequent intervals in the Chinese hamster dihydrofolate reductase origin of replication but are used with very different efficiencies. Mol Cell Biol 22:3053–3065.

Ferguson BM, Fangman WL. 1992. A position effect on the time of replication origin activation in yeast. Cell 68:333–339.

Ghosh M, Liu G, Randall G, Bevington J, Leffak M. 2004. Transcription factor binding and induced transcription alter chromosomal c-myc replicator activity. Mol Cell Biol 24:10193–10207.

Giacca M, Zentilin L, Norio P, Diviacco S, Dimitrova D, Contreas G, Biamonti G, Perini G, Weighardt F, Riva S. 1994. Fine mapping of a replication origin of human DNA. Proc Natl Acad Sci USA 91:7119–7123.

Gilbert DM. 2004. In search of the holy replicator. Nat Rev Mol Cell Biol 5:848–855.

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

Gomez M, Antequera F. 1999. Organization of DNA replication origins in the fission yeast genome. EMBO J 18:5683–5690.

Gray SJ, Liu G, Altman AL, Small LE, Fanning E. 2007. Discrete functional elements required for initiation activity of the Chinese hamster dihydrofolate reductase origin beta at ectopic chromosomal sites. Exp Cell Res 313:109–120.

Haase SB, Heinzel SS, Calos MP. 1994. Transcription inhibits the replication of autonomously replicating plasmids in human cells. Mol Cell Biol 14:2516–2524.

Hamlin JL. 1992. Mammalian origins of replication. Bioessays 14:651–659.

Handeli S, Klar A, Meuth M, Cedar H. 1989. Mapping replication units in animal cells. Cell 57:909–920.

Harland RM, Laskey RA. 1980. Regulated replication of DNA microinjected into eggs of Xenopus laevis. Cell 20:761–771.

Heck MM, Spradling AC. 1990. Multiple replication origins are used during Drosophila chorion gene amplification. J Cell Biol 110:903–914.

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

Huberman JA, Riggs AD. 1968. On the mechanism of DNA replication in mammalian chromosomes. J Mol Biol 32:327–341.

Hyrien O, Maric C, Mechali M. 1995. Transition in specification of embryonic metazoan DNA replication origins. Science 270:994–997.

Jacob F, Brenner S. 1963. Sur la Regulation de la synthese du DNA chez les bacteriers: l'hypothese du replicon. Comp Rendu Acad Sci 246:298–300.

Jeon Y, Bekiranov S, Karnani N, Kapranov P, Ghosh S, MacAlpine D, Lee C, Hwang DS, Gingeras TR, Dutta A. 2005. Temporal profile of replication of human chromosomes. Proc Natl Acad Sci USA 102:6419–6424.

Kalejta RF, Li X, Mesner LD, Dijkwel PA, Lin HB, Hamlin JL. 1998. Distal sequences, but not ori-beta/OBR-1, are essential for initiation of DNA replication in the Chinese hamster DHFR origin. Mol Cell 2:797–806.

Keezer SM, Gilbert DM. 2002. Sensitivity of the origin decision point to specific inhibitors of cellular signaling and metabolism. Exp Cell Res 273:54–64.

Kim SM, Huberman JA. 2001. Regulation of replication timing in fission yeast. EMBO J 20:6115–6126.

Kitsberg D, Selig S, Keshet I, Cedar H. 1993. Replication structure of the human beta-globin gene domain. Nature 366:588–590.

Krysan PJ, Haase SB, Calos MP. 1989. Isolation of human sequences that replicate autonomously in human cells. Mol Cell Biol 9:1026–1033.

Lande-Diner L, Cedar H. 2005. Silence of the genes—Mechanisms of long-term repression. Nat Rev Genet 6:648–654.

Larner JM, Lee H, Little RD, Dijkwel PA, Schildkraut CL, Hamlin JL. 1999. Radiation down-regulates replication origin activity throughout the S phase in mammalian cells. Nucl Acids Res 27:803–809.

Leu TH, Hamlin JL. 1989. High-resolution mapping of replication fork movement through the amplified dihydrofolate reductase domain in CHO cells by in-gel renaturation analysis. Mol Cell Biol 9:523–531.

Lin HB, Dijkwel PA, Hamlin JL. 2005. Promiscuous initiation on mammalian chromosomal DNA templates and its possible suppression by transcription. Exp Cell Res 308:53–64.

Liu G, Malott M, Leffak M. 2003. Multiple functional elements comprise a Mammalian chromosomal replicator. Mol Cell Biol 23:1832–1842.

MacAlpine DM, Bell SP. 2005. A genomic view of eukaryotic DNA replication. Chromosome Res 13:309–326.

MacAlpine DM, Rodriguez HK, Bell SP. 2004. Coordination of replication and transcription along a Drosophila chromosome. Genes Dev 18:3094–3105.

Mesner LD, Hamlin JL. 2005. Specific signals at the 3′ end of the DHFR gene define one boundary of the downstream origin of replication. Genes Dev 19:1053–1066.

Mesner LD, Hamlin JL, Dijkwel PA. 2003a. The matrix attachment region in the Chinese hamster dihydrofolate reductase origin of replication may be required for local chromatid separation. Proc Natl Acad Sci USA 100:3281–3286.

Mesner LD, Li X, Dijkwel PA, Hamlin JL. 2003b. The dihydrofolate reductase origin of replication does not contain any nonredundant genetic elements required for origin activity. Mol Cell Biol 23:804–814.

Mesner LD, Crawford EL, Hamlin JL. 2006. Isolating apparently pure libraries of replication origins from complex genomes. Mol Cell 21:719–726.

Nawotka KA, Huberman JA. 1988. Two-dimensional gel electrophoretic method for mapping DNA replicons. Mol Cell Biol 8:1408–1413.

Newlon CS, Petes TD, Hereford LM, Fangman WL. 1974. Replication of yeast chromosomal DNA. Nature 247:32–35.

Norio P, Kosiyatrakul S, Yang Q, Guan Z, Brown NM, Thomas S, Riblet R, Schildkraut CL. 2005. Progressive activation of DNA replication initiation in large domains of the immunoglobulin heavy chain locus during B cell development. Mol Cell 20:575–587.

Paixao S, Colaluca IN, Cubells M, Peverali FA, Destro A, Giadrossi S, Giacca M, Falaschi A, Riva S, Biamonti G. 2004. Modular structure of the human lamin B2 replicator. Mol Cell Biol 24:2958–2967.

Pasero P, Bensimon A, Schwob E. 2002. Single-molecule analysis reveals clustering and epigenetic regulation of replication origins at the yeast rDNA locus. Genes Dev 16:2479–2484.

Pattenden SG, Chandy MJ, Gutierrez JL, Workman JL. 2005. Chromatin dynamics rule the genome. Genome Biol 6:355.

Raghuraman MK, Winzeler EA, Collingwood D, Hunt S, Wodicka L, Conway A, Lockhart DJ, Davis RW, Brewer BJ, Fangman WL. 2001. Replication dynamics of the yeast genome. Science 294:115–121.

Razin SV, Kekelidze MG, Lukanidin EM, Scherrer K, Georgiev GP. 1986. Replication origins are attached to the nuclear skeleton. Nucl Acids Res 14:8189–8207.

Saha S, Shan Y, Mesner LD, Hamlin JL. 2004. The promoter of the CHO dihydrofolate reductase gene regulates the activity of the local origin and helps define its boundaries. Genes Dev 18:397–410.

Schubeler D, MacAlpine DM, Scalzo D, Wirbelauer C, Kooperberg C, van Leeuwen F, Gottschling DE, O'neill LP, Turner BM, Delrow J, Bell SP, Groudine M. 2004. The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. Genes Dev 18:1263–1271.

Segurado M, de Luis A, Antequera F. 2003. Genome-wide distribution of DNA replication origins at A+T-rich islands in Schizosaccharomyces pombe. EMBO Rep 4:1048–1053.

Snyder M, Sapolsky RJ, Davis RW. 1988. Transcription interferes with elements important for chromosome maintenance in Saccharomyces cerevisiae. Mol Cell Biol 8:2184–2194.

Stinchcomb DT, Struhl K, Davis RW. 1979. Isolation and characterisation of a yeast chromosomal replicator. Nature 282:39–43.

Todorovic V, Giadrossi S, Pelizon C, Mendoza-Maldonado R, Masai H, Giacca M. 2005. Human origins of DNA replication selected from a library of nascent DNA. Mol Cell 19:567–575.

Toledo F, Baron B, Fernandez MA, Lachages AM, Mayau V, Buttin G, Debatisse M. 1998. oriGNA I3: A narrow zone of preferential replication initiation in mammalian cells identified by 2D gel and competitive PCR replicon mapping techniques. Nucl Acids Res 26:2313–2321.

Vogelauer M, Rubbi L, Lucas I, Brewer BJ, Grunstein M. 2002. Histone acetylation regulates the time of replication origin firing. Mol Cell 10:1223–1233.

Wright AP, Maundrell K, Shall S. 1986. Transformation of Schizosaccharomyces pombe by non-homologous, unstable integration of plasmids in the genome. Curr Genet 10:503–508.

Zhang J, Xu F, Hashimshony T, Keshet I, Cedar H. 2002. Establishment of transcriptional competence in early and late S phase. Nature 420:198–202.

Zhou J, Chau C, Deng Z, Stedman W, Lieberman PM. 2005. Epigenetic control of replication origins. Cell Cycle 4:889–892.

Zhou J, Chau CM, Deng Z, Shiekhattar R, Spindler MP, Schepers A, Lieberman PM. 2006. Cell cycle regulation of chromatin at an origin of DNA replication. EMBO J 24:1406–1417.