Molecular Biology of the Cell
Vol. 10, 757–769, March 1999

Short DNA Fragments without Sequence Similarity Are Initiation Sites for Replication in the Chromosome of the Yeast *Yarrowia lipolytica*

Laurence Vernis,*† Marion Chasles,* Philippe Pasero,‡§ Andrée Lepingle,* Claude Gaillardin,* and Philippe Fournier*‖

*Laboratoire de Génétique Moléculaire et Cellulaire, Institut National de la Recherche Agronomique-Centre National de la Recherche Scientifique, 78850 Thiverval-Grignon, France; and ‡Laboratoire de Génétique, Centre National de la Recherche Scientifique, Faculté de Médecine, 13385 Marseille Cedex 5, France

Submitted June 10, 1998; Accepted December 7, 1998
Monitoring Editor: David Botstein

We have previously shown that both a centromere (CEN) and a replication origin are necessary for plasmid maintenance in the yeast *Yarrowia lipolytica* (Vernis *et al.*, 1997). Because of this requirement, only a small number of centromere-proximal replication origins have been isolated from *Yarrowia*. We used a *CEN*-based plasmid to obtain noncentromeric origins, and several new fragments, some unique and some repetitive sequences, were isolated. Some of them were analyzed by two-dimensional gel electrophoresis and correspond to actual sites of initiation (*ORI*) on the chromosome. We observed that a 125-bp fragment is sufficient for a functional *ORI* on plasmid, and that chromosomal origins moved to ectopic sites on the chromosome continue to act as initiation sites. These *Yarrowia* origins share an 8-bp motif, which is not essential for origin function on plasmids. The *Yarrowia* origins do not display any obvious common structural features, like bent DNA or DNA unwinding elements, generally present at or near eukaryotic replication origins. *Y. lipolytica* origins thus share features of those in the unicellular *Saccharomyces cerevisiae* and in multicellular eukaryotes: they are discrete and short genetic elements without sequence similarity.

INTRODUCTION

The nature of the initiation sites for DNA replication has been investigated in many different eukaryotes, and a general model, largely based on viral systems, has been proposed (DePamphilis, 1996). This model defines an origin of replication as a core element flanked by a DNA unwinding element (DUE) and nonessential auxiliary sequences (e.g., binding sites for transcription factors); however, several origins from various organisms do not conform to this general scheme. The overall sizes of replication origins are diverse: several base pairs in some viral genomes, a few hundred base pairs in unicellular organisms such as *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, or *Physarum polycephalum* (Maundrell *et al.*, 1988; Newlon and Theis, 1993; Bénard *et al.*, 1996; Dubey *et al.*, 1996), and several kilobase pairs in some mammalian loci, such as the dihydrofolate reductase region of hamster cells or the human *ADA* gene (Vita-Pearlman *et al.*, 1993; Dijkwel *et al.*, 1994). The nature and the relative positions of these sequences also differ significantly from one organism to another. The core element, for example, is a short consensus sequence in *S. cerevisiae* (the 11-bp autonomously replicating sequence [*ARS*] consensus sequence [ACS]; Newlon and Theis, 1993). It is composed of one or several essential stretches of 30–50 bp in *Kluyveromyces lactis* (Fabiani *et*

† Corresponding author. E-mail address: vernis@platon.grignon.inra.fr.
Present addresses: §Institut de Génétique Moléculaire, Centre National de la Recherche Scientifique, 34033 Montpellier, France; ‖Institut National de la Recherche Agronomique-Centre National de la Recherche Scientifique, 30380 St.-Christol-les-Alès, France.

al., 1996) or in *S. pombe* (Dubey *et al.*, 1996). In other cases, for example in *P. polycephalum*, it is a quite large initiation zone without any obvious conserved motifs (Bénard *et al.*, 1996). The term core element is therefore inappropriate in these cases, and perhaps also in systems where initiation itself occurs at numerous dispersed sites (Dijkwel *et al.*, 1994; Shinomiya and Ina, 1994; Dhar *et al.*, 1996).

Current models for initiation of DNA replication stipulate that an initiator protein binds the origin to promote initiation of DNA replication (for review, see Diffley, 1996). The origin recognition complex (ORC) in the budding yeast *S. cerevisiae* is a complex of six polypeptides that binds the ACS and the adjacent B1 element (Lee and Bell, 1997, and references therein). This complex is always present at the replication origins (Diffley *et al.*, 1994), which implies that steps other than ORC binding must be cell cycle-regulated (Diffley *et al.*, 1995; Rowley *et al.*, 1995). In vivo foot-printing studies have shown that the chromatin structure at the *ARS* elements changes during the cell cycle (Diffley *et al.*, 1994). A so-called prereplicative complex assembles in G_1, with the loading of MCMs by Cdc6p at the origins (Tanaka *et al.*, 1997), and initiation is triggered by cyclin-dependent kinases at the G_1/S boundary (Jallepalli and Kelly, 1997). All of these initiation factors identified in yeast have highly conserved homologues in all other eukaryotes, from *K. lactis* and *S. pombe* to the multicellular organisms *Arabidopsis*, *Drosophila*, *Xenopus*, *Caenorhabditis*, and *Homo sapiens* (for review, see Diffley, 1996). In many organisms, these proteins have been shown to be essential for DNA synthesis, suggesting the existence of a common mechanism controlling the initiation of DNA replication (Diffley, 1996).

In *Xenopus* egg extracts, ORC, Cdc6p, and the MCMs regulate initiation of DNA replication at apparently random sites (Rowles and Blow, 1997). In yeast, the same factors are involved in the regulation of precisely defined origins. It is unclear how conserved *trans*-acting factors interact with such a variety of *cis*-acting sequences. The study of simple eukaryotes with replication origins structurally different from those of *S. cerevisiae* may help elucidate what defines a chromosomal origin of DNA replication. In *S. pombe*, for example, replication origins consist of 600- to 800-bp sequences containing several essential regions, which include AT-rich stretches (Zhu *et al.*, 1994; Dubey *et al.*, 1996). Interference has been shown between very close origins identified within the *ura4* chromosomal locus (Dubey *et al.*, 1994), whereas a plasmid harboring any of these origins replicates autonomously in that yeast. On the other hand, discrete initiation sites have also been mapped on *S. pombe* chromosomes, like those in *S. cerevisiae* (Caddle and Calos, 1994; Wohlgemuth *et al.*, 1994). In *P. polycephalum* the origins identified are located within the ribosomal DNA repeated unit (Bénard *et al.*, 1995) or at promoter regions of actively transcribed genes (Bénard *et al.*, 1996). They are not AT rich and do not confer autonomous replication to plasmids. In the yeast *Yarrowia lipolytica*, *ARS*s are selected on plasmids constructed from a genomic library at a very low frequency, i.e.,. one every 3 Mb (Fournier *et al.*, 1991; Matsuoka *et al.*, 1993). These replicative plasmids are maintained at low copy number (one to three per cell) with a relatively high mitotic stability (3% loss per cell generation). Genetic studies in meiosis and analysis of integrative transformation events for two such sequences indicate that they each contained a centromere (Fournier *et al.*, 1993; Vernis *et al.*, 1997); however, the replication does not start at the centromere, and an adjacent site of initiation was demonstrated in the cloned *ARS*. The centromere and the origin of replication can be physically separated on the plasmid, and the origin is also active and essential for initiation in its chromosomal locus (Vernis *et al.*, 1997). These observations suggested that it should be possible to clone chromosomal origins using a centromeric vector. We report here the detailed analysis of the replication origins obtained by this strategy of screening randomly generated genomic fragments. We show that initiation sites for DNA replication are present around every 20- to 50-kb in the *Y. lipolytica* genome. These sites are intermediate in structure between the short and simple *ARS* elements of *S. cerevisiae* and the apparently nonconserved origin regions of other eukaryotes.

MATERIALS AND METHODS

Strains

Escherichia coli strains used (DK1 and XL1-Blue-MRF′) have been described previously (Vernis *et al.*, 1997). The *Y. lipolytica* recipient strain used for transformation was INAG33122 (*MatB*, *lys2*-5, *leu2*-35, *ade1*, *xpr2*), the reference strain for chromosome separation was E150 (*MatB*, *his*-1, *leu2*-270, *ura3*-302, *xpr2*-322), and the wild-type strain W29 was used to prepare genomic DNA for the gene bank. All three strains are deposited at the Collection de Levures d'Intérêt Biotechnologique (Grignon) under the numbers CLIB118, CLIB122, and CLIB89, respectively. Culture media were as described by Barth and Gaillardin (1996). Mitotic stability of plasmids was measured as described by Vernis *et al.* (1997).

Molecular Techniques

Y. lipolytica was transformed as described in Barth and Gaillardin (1996). Chromosomes were separated by published protocols (Zimmermann and Fournier, 1996). The methods for preparation of replication intermediates and two-dimensional (2D) gel analyses were those described by Huberman (1994). Standard techniques of DNA manipulation in *E. coli* and DNA hybridization were used (Sambrook *et al.*, 1989).

Mutations in *ori*1068 and *ori*3018 were introduced by PCR amplification using primers harboring the mutation (ATCGATAA instead of TACAAGTA) and ended by an *Sal*I site. A three-way PCR was needed in the case of *ori*3018, using a third central primer harboring

the mutation, because the consensus was located in the central part of the origin.

An artificial I-*Sce*I was introduced at the *Cla*I site of pINA989 obtained from a DNA library (Xuan *et al.*, 1988) in *E. coli* by colony hybridization. pINA989 contains *ori*X009 on a 5.5-kb genomic fragment, and the *LEU2* gene. After linearization with *Bam*HI in the *ori*X009 region, this construct was introduced by targeted integration into the *Yarrowia* genome.

DNA Library

The construction of pINA732 used for cloning origins of replication was described by Vernis *et al.*, 1997. We used a *Bam*HI deletion of this plasmid, called pINA732-Δ. *Sau*3a-digested genomic DNA was ligated into the *Bam*HI site of pINA732-Δ. A set of 5700 *E. coli* transformants was obtained. More than 90% of the plasmids carried genomic inserts, with a mean size of 1 kb. To check the size of the cloned inserts in yeast, we used two primers: 305 is a 17-bp oligonucleotide corresponding to the sequence starting at base 305 of pBR322, and 2216 is 28 bp corresponding to the sequence starting at base 2216 in the sequence of *ARS*68 (GenBank M91601).

The size of the inserts was determined to plasmids in 53 yeast transformants, and 11 short sequences were selected (<600 bp) for further analysis. Only 7 of these 11 were different, because one (the genomic repeated *ori*X009 sequence; see below) was found five times. Total DNA of only five of these seven yeast transformants was able to transform *E. coli*.

Colinearity with yeast genome was checked by PCR amplification of plasmid DNA, total DNA from a yeast transformant, and wild-type genomic DNA. A band of identical size was seen in all three cases for every cloned origin, showing that no major rearrangement had occurred during the cloning experiment (our unpublished results).

Sequences of PCR amplification products were determined after purification of the DNA with a PCR Purification Kit (Qiagen, Hilden, Germany).

DNA Sequence Analysis

Sequences were compared using the GCG package (University of Wisconsin, Madison, WI) or the Macaw software (G. Schuler, National Center for Biotechnology Information, Bethesda, MD) using the Gibbs algorithm. The 3D path of DNA molecules was calculated using the algorithm of Eckdahl and Anderson (1987) and the helical parameters of Bolshoy *et al.* (1991), as described previously (Pasero *et al.*, 1993). The magnitude of DNA bending on curvature maps is expressed as the ENDS ratio, which is defined as the ratio of the contour length of a segment of the helical axis to the shortest distance between its ends. ENDS ratios were computed for a window of 120 nucleotides and a step of 10 nucleotides. The thermodynamic library of Breslauer *et al.* (1986), characterizing the ten Watson–Crick nearest-neighbor interactions in DNA, was used to predict the stability of DNA duplexes as described previously by Kowalski and coworkers (Natale *et al.*, 1993). The Δ*G* values presented in Figure 6 were calculated for a 120-bp DNA duplex in 1 M NaCl, pH 7.0, at 25°C. Note that the parameters used in the Thermodyn program of Kowalski are slightly different and lead to values that are proportionally lower. Variations of Roll angle and AT% were calculated as described previously (Marilley and Pasero, 1996) for a 120-bp window and a 10-bp step. The distribution of AT-rich sequences within origins was analyzed using a program developed by P. Trécourt (Mathematics Department, Institut National Agronomique-Paris-Grignon, 75231 Paris Cedex 05, France). The algorithm generates random sequences of given length and base composition and can be used to score the occurrence of a given motif. The results are saved in a format that can be exploited by the Excel software. Lines of the program are as follows (separated by//): dim a%(20)//open ≪0≫; #1, ≪dnaseq.txt≫//input ≪number of tests?≫,nsim%//input ≪length of the sequence?≫,nls//input ≪number of A and T?≫,no//if no >= nls then end//freq=no/nls//temp%=0:randomize(timer)//for l = 1 to nsim%//for i = 1 to 20 :a%(i)=0:next i//for i = 1 to nls//x=rnd(1)//if x < freq then incr temp% else fin%=1//if fin%=1 then//if temp% > 0 then//if temp% > 20 then temp%=20//incr a%(temp%)//end if//temp%=0//fin%=0//end if//next i//if temp% > 0 the, incr a%(temp%)//for i = 1 to 19//print using ≪###≫;a%(i)//print#1,using ≪###≫;a%(i)//next i//print using ≪ ##≫;a%(20)//print#1,using ≪ ##≫;a%(20)//print//next l//print ≪results are saved in the file DNASEQ.TXT≫//end.

RESULTS

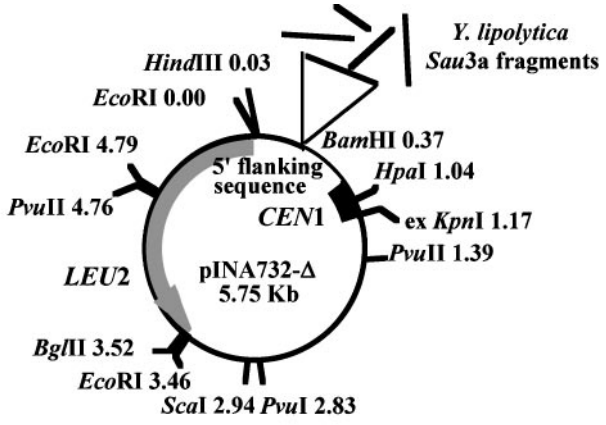
Cloning Origins of Replication with a Centromeric Plasmid

Replicative vectors in *Y. lipolytica* require both a centromere and a replication origin. Genomic DNA was ligated into a *CEN1*-based plasmid (see MATERIAL AND METHODS) (Figure 1), and the resulting library was used to transform *Y. lipolytica* by electroporation. More than 90% of the transformants displayed mitotic instability. Five plasmids carrying small-sized inserts were back-transformed into *E. coli* and further analyzed. The colinearity of the fragments with genomic DNA was checked by PCR. Purified plasmid DNA transformed *Y. lipolytica* at high frequency (replicative transformants), showing that the inserts confer autonomous replication to a centromeric plasmid (Figure 1). From the transformation frequency of the plasmid pool as well as of individual cloned *ARS*s, we estimate that there is one putative origin of replication for every 20–50 kb of the *Y. lipolytica* genome.

The Cloned Sequences Represent a Wide Variety of Genomic Loci

Several types of ribosomal DNA repeats have been described in *Y. lipolytica* that differ in the sequence of their nontranscribed spacers (NTSs). Two of them have previously been cloned and sequenced (G and P2 units) (van Heerikhuizen *et al.*, 1985). One of the clones obtained here in our shotgun cloning of origins displays an almost perfect sequence identity with the common part of the two known NTSs. Because rDNA is a repeated sequence, it should be found several times in the gene bank, but our selection focused on smaller DNA pieces, so that only one such clone was analyzed. It was called *ori*-rDNA. The fragment isolated was 597 bp, and a functional 285-bp subfragment was obtained by a *Hin*dIII deletion between a site in the pBR322 vector and an internal site. A putative replication origin is thus present in the *Y. lipolytica* NTSs as in the yeasts *S. cerevisiae* and *S. pombe* (Linskens and Huberman, 1988; Sanchez *et al.*, 1998a), the slime mold *P. polycephalum* (Bénard *et al.*, 1995), and many other eukaryotes. *Ori*-rDNA hybridizes to the five chromosomal bands on an electrophoretic karyotype (Figure 1), which is consistent with the

A



B

	chromosomal band	size	AT%	chromosomal activity (2D)
centromeric origins				
*ori*4002	IV	789 bp	54,1%	nt
*ori*3018*	III	144 bp	60,4%	active
*ori*1068*	I	125 bp	57,6%	active
non-centromeric origins				
repeated sequences				
*ori*X009	I to V	158 bp	61,0%	nt
*ori*X096	III, IV, V	564 bp	51,1%	nt
ori-rDNA	I to V	285 bp	60,9%	active
unique sequences				
*ori*3068	III	115 bp	62,6%	nt
*ori*4021	IV	393 bp	52,2%	active

Figure 1. Cloning of *Y. lipolytica* origins with a centromeric plasmid. (A) The cloning vector pINA732-Δ contains the *LEU2* gene (dotted area), the *CEN1* centromere, and flanking genomic sequences. *Y. lipolytica* genomic DNA digested with *Sau3A* was ligated into the *Bam*HI site of the vector. (B) Characteristics of the eight *ORI* elements analyzed in this article. Demonstration that five of them are active origins comes from Vernis *et al.* (1997) and from data presented below in this article. We therefore adopt the *ORI* designation for all of them. Two of these sequences, marked by a star (*ori*1068 and *ori*3018), were subjected to systematic size reduction to determine the minimal functional fragment (Figure 4). The chromosomal localization of the different fragments was determined by pulsed-field gel electrophoresis and Southern blotting using reference strain E150. Chromosomal bands are numbered according to Vernis *et al.* (1997). Chromosomal activity (active; nt, not tested) refers to the observation of replication initiation signals after 2D gel electrophoresis.

presence of several independent clusters of rDNA in the *Y. lipolytica* genome (Fournier *et al.*, 1986), and with their mapping on several chromosomes in different strains (Casarégola *et al.*, 1997).

Two other sequences (*ori*X009 and *ori*X096) hybridize to several chromosomal bands (Figure 1). Further genomic mapping by Southern hybridization suggested that each of these sequences is part of a larger element and that both elements are present in multiple copies. We compared the sequences of *ori*X009 and *ori*X096 with the *Ylt1* transposon of *Y. lipolytica* (Schmid-Berger *et al.*, 1994) and found no similarity. We also compared the genomic maps of *ori*X009 and *ori*X096 loci with the known maps of rDNA (van Heerikhuizen *et al.*, 1985) and mit-DNA (Wesolowski *et al.*, 1981); they did not appear to originate from either of these repeated DNAs. We then investigated whether *ori*X009 is a subtelomeric sequence. Indeed, it has been reported that an *ARS*, although generally inactive in the chromosome, is closely associated with the telomere in *S. cerevisiae* (Newlon *et al.*, 1993). We introduced a rare cutting I-*Sce*I site near *ori*X009 sequences in the chromosomes to measure the distance between *ori*X009 and the chromosome end after an I-*Sce*I cut. pINA789 containing 5.5 kb flanking *ori*X009 was isolated from a genomic library and tagged with an I-*Sce*I site (see MATERIALS AND METHODS). This plasmid was linearized by *Bam*HI within the *ori*X009 flanking sequences and integrated into one *ori*X009 region by integrative transformation. The structure of the integrated DNA was checked by Southern blotting and hybridization. Chromosomal plugs of several transformants were prepared in agarose, digested with I-*Sce*I, and separated by field inversion gel electrophoresis. The gel was blotted and hybridized with a pBR322 probe, which revealed bands ranging from 20 to 500 kb (Figure 2). This indicates that *ori*X009 sequences are located much further away from chromosome ends than are subtelomeric repeats in *S. cerevisiae*.

*Ori*3068 and *ori*4021 each hybridized to a single chromosomal band. Genomic restriction digestion and Southern hybridization experiments indicated that these were only single copies of these origins in the genome. A restriction map of each region was established, which confirmed the colinearity between the cloned fragment and the genomic locus.

Various genomic regions, either repetitive (*ori*-rDNA, *ori*X009, *ori*X096) or unique (*ori*3068, *ori*4021), from different chromosomes are thus able to confer extrachromosomal maintenance to a *CEN1* plasmid. This confirms our previous interpretation of our initial *ARS* assay (Vernis *et al.*, 1997): only centromere-proximal origins were then cloned, not because of some specific features displayed by these sequences but because of the absolute requirement for a centromere to maintain plasmids in *Y. lipolytica*.

Ori4021 Is Active in the Chromosome

To test the chromosomal activity of these sequences, replication intermediates were prepared and separated on a 2D gel and then hybridized with an origin probe (see MATERIALS AND METHODS). In the case of the single copy *ori*4021 sequence, a 5.7-kb *Bam*HI restriction fragment centered around the putative or-

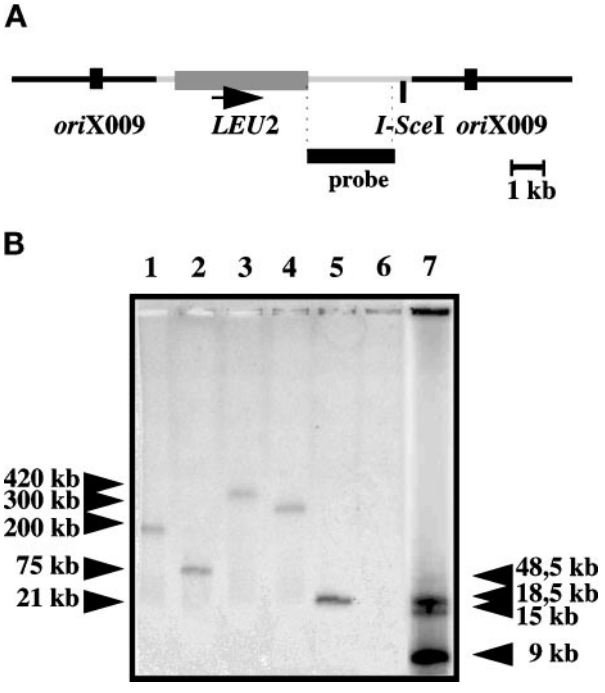


Figure 2. Localization of the repeated *ori*X009 sequence in the genome of *Y. lipolytica*. (A) Map of a chromosomal integration at an *ori*X009 locus of the pINA989 containing *ori*X009 within its genomic environment, the *LEU2* gene, and an I-*Sce*I site. pBR322 vector sequences are shown in light gray. (B) Chromosomes from various transformants (lanes 1–5) were embedded in agarose plugs, digested with I-*Sce*I, and separated under the following field inversion gel electrophoresis conditions: 1% pulsed field agarose (Bio-Rad, Bio-Rad, Hercules, CA) in 0.5× TAE buffer, run at 6 volts/cm for 14 h at 14°C with an angle of 106° and a variation of pulse from 0.41 sec to 2 min 6.52 sec. The DNA was transferred to a membrane and hybridized with a pBR322 probe (black box). A nontransformed strain is shown as a control in lane 6. The Raoul marker (Appligene, Heidelberg, Germany) is in lane 7. *S. cerevisiae* chromosomes and λ concatemers from Bio-Rad (not visible on the autoradiogram) were also used as molecular weight markers. As a comparison, the smallest chromosome among the six from *Yarrowia* is 2.6 Mb, and the largest is 4.9 Mb.

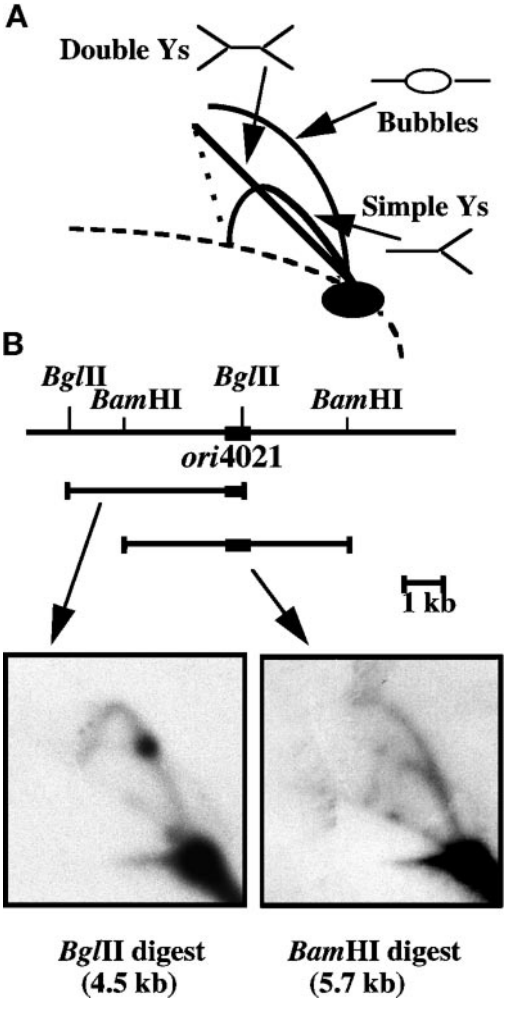


Figure 3. *Ori*4021 is active in the genome. (A) Sketch showing the migration patterns of various types of replication intermediates during neutral/neutral 2D gel electrophoresis. B) A restriction map of the genomic locus shows the two *Bam*HI and *Bgl*II restriction fragments that are separated on a 2D gel and hybridized with an *ori*4021 probe. The pause signal observed along the Y arc on the *Bgl*II digest is not discussed here.

igin was analyzed (Figure 3). A typical bubble arc was visible on the 2D gel, indicative of an initiation in the central third of the restriction fragment. A faint Y signal, which may occur if the origin fails to fire at every cell cycle, was also detected; however, this Y arc is unusual because large Y-shaped intermediates are less visible. Possibly this was due to 1) the superposition of two Y arcs resulting from partial digest with *Bam*HI or as 2) a diffuse termination signal in this zone as shown on Figure 3A. On a *Bgl*II restriction fragment (control), only the Y arc was present, confirming that the *ori*4021 initiation site corresponds to a discrete locus and is not part of a large initiation zone, as is the case for some higher eukaryote loci (for review, see DePamphilis, 1996). An accumulation of molecules is also visible as a spot on the ascending part of the Y arc, which may indicate the presence of a replication pause site in that region.

On the basis of existing restriction maps of ribosomal DNA units (van Heerikhuizen *et al.*, 1985), a 2D gel analysis of *Eco*RI-digested genomic DNA, using *ori*-rDNA as a probe, was expected to reveal *ori*-rDNA activity. The pattern obtained was actually complex, with degradation products masking some of the replication intermediates; however, at least two bubble arcs indicative of initiation sites were detected, as well as Y arcs (our unpublished results). This complex pattern may reflect a length polymorphism of the NTS and possibly initiation at only a subset of the rDNA units, as is the case for *S. cerevisiae* (Linskens and Huberman, 1988).

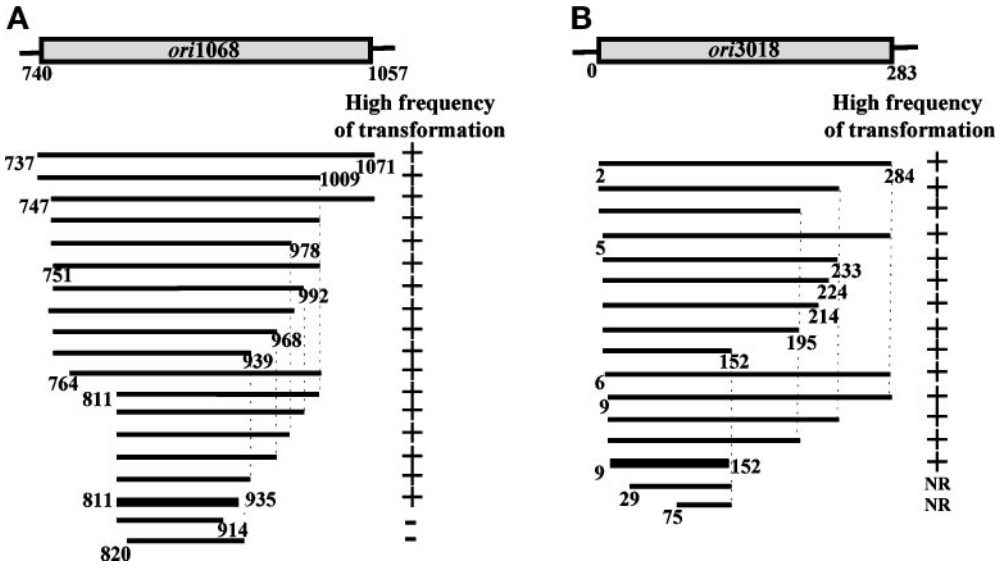


Figure 4. Mapping the minimal functional fragment of *ori*1068 (A) and *ori*3018 (B). Position and orientation of the synthetic primers are shown on the top line. Nucleotides positions are according to GenBank M91601 and M91600. The strain INAG3122 was transformed by electroporation with centromeric plasmids containing the reduced origins, because only replicative transformation is obtained at high frequency with this method. Results of transformation are shown by + and −. NR, Not recovered in *E. coli*.

Thus *ori*4021 and *ori*-rDNA confirm that the use of a centromeric vector to clone genomic DNA allows the isolation of active chromosomal origins of replication. The centromeric origins described previously (Vernis *et al.*, 1997) are therefore not the only replication loci on the chromosome.

The Origin Size Can Be Reduced to 125 bp on Plasmid

To define more precisely what is the minimal size required for initiating replication, we focused on the two previously characterized centromeric origins *ori*3018 and *ori*1068, which are active both on plasmids and in the chromosome location (Vernis *et al.*, 1997). Fragments of different lengths were synthesized by PCR amplification (Figure 4) and inserted into the *Sal*I site of the *LEU*2-*CEN*1 vector pINA732 (Fournier *et al.*, 1993). The resulting plasmids were checked by digestion and by sequencing the origin and were subsequently used to transform *Y. lipolytica. Ori*1068 reduced to a 125-bp fragment and *ori*3018 reduced to a 144-bp fragment are each sufficient to transform yeast at high frequency (Figure 4). It is possible that *ori*3018 could be reduced further, but smaller amplification fragments were structurally unstable in *E. coli*. No significant changes in transformation efficiency were observed between the various deletions, so no functionally essential sequences appeared to map outside the minimal origins. We also checked that the yeast transformants obtained with the smallest origin sequences were still mitotically unstable like the initial *ORI* plasmids, and that the plasmids were present as monomeric CCC molecules. The small length of these deleted centromeric origins is comparable to the size of other genomic origins cloned above, like *ori*3068 (115 bp) or *ori*X009 (159 bp), and we turned to an extensive sequence analysis of these regions.

Six Yarrowia Origins Share an 8-bp Motif That Is Dispensable for ARS Function

We looked for the presence of the *S. cerevisiae* 11-bp *ARS* consensus sequence (WTTTAYRTTTW) in the eight *Yarrowia* origins and found no perfect match. Several but not all origins displayed 9/11 and 10/11 matches. Neither the 40-bp core sequence of *K. lactis ARS* (Fabiani *et al.*, 1996) nor motifs frequently found in *S. pombe* origins (Maundrell *et al.*, 1988, Dubey *et al.*, 1996) were present. We therefore looked for a sequence common to the *Y. lipolytica* origins. Several sequence alignment programs were used (see MATERIALS AND METHODS), and we repeatedly found copies of a short 8-bp motif TDCAAGTH (D = A or G or T; H = A or C or T), which is present one to five times in all the origins, except in *ori*X009 and in *ori*X096 (Figure 5). To assess the role of this sequence in origin function, we mutated the TACAAGTA sequence in *ori*1068 and *ori*3018 to redistribute the bases within this 8-bp stretch without affecting the overall base composition. In *ori*3018, the two partially overlapping motifs were both destroyed. The resulting sequences ligated into the centromeric vector still transformed *Y. lipolytica* with high frequency. The consensus is therefore not essential for origin function on a plasmid.

Because *Yarrowia* origins apparently do not share any essential sequence consensus, we then looked for the overrepresentation of short motifs. Indeed, Zhu *et al.* (1994) have observed that some AT-rich hexanucleotides are abundant in *S. pombe ARS*, suggesting that they could be important for initiation. To perform a

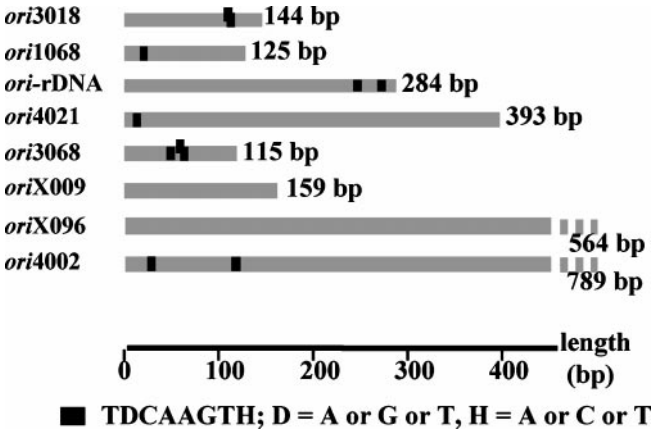


Figure 5. Localization of the consensus sequence TDCAAGTH in six of the eight *Y. lipolytica* origins. Nucleotide sequences are deposited under the following GenBank accession numbers: AF038941 (*ori*X009), AF038942 (*ori*X096), AF038943 (*ori*4021), AF039581 (*ori*-rDNA), and AF038940 (*ori*3068).

statistical analysis of the distribution of nucleotides in *Yarrowia* origins, we used a program developed in our lab (see MATERIALS AND METHODS) that generates random sequences of a given length and a given composition. Five thousand random DNA sequences of the same length and base composition as each *Yarrowia* origin were generated, and the occurrence of AT-rich oligonucleotides from SW_1S to $SW_{20}S$ (where S stands for G or C) was scored. The frequency of occurrence of these motifs in a *Yarrowia* origin was then compared with the frequency of occurrence of the same motifs in the generated sequences using a χ^2 test. We failed to detect any significantly unusual distribution of oligonucleotides in the *Yarrowia* origins. As a control, the same program was applied to scaffold-attached regions (SARs), which have been reported to be frequently associated with origins of replication (DePamphilis, 1996) and are characterized by a biased distribution of AT repeats (Roberge and Gasser, 1992). With the *Drosophila fushi tarazu* (*ftz*) SAR or with the SAR associated with the histone gene, a significant divergence from a random distribution was detected by the program (χ^2 20.3 vs. 15.51 in a standard χ^2 table for *ftz*, and 42.77 vs. 18.31 for *HIST*). We conclude that the eight *Yarrowia* origins identified so far do not contain any significant reiteration of primary sequence motifs.

We also used the same program to analyze two origins from higher eukaryotes: the 55% AT-rich 500 bp of the *ori*β locus in the hamster DHFR origin region (Burhans *et al.*, 1990), and the 51% AT-rich 1994 bp of the origin associated with the mouse *ADA* gene (Vita-Pearlman *et al.*, 1993). Both gave a significant χ^2 value (17.87 vs. 11.07 in the first case, and 47.07 vs. 15.51 in the second), the bias being due to an excess of the classes SWS and SWWS and to a lower representation of classes with >4 W. This result is in accordance with the presence of alternating purines, which has been described in these origins as well as in other eukaryotic origins (Bergemann and Johnson, 1992). A protein factor displays affinity for the purine motif GGNNGAGGGAGARRRR (R = purine; N = any base), but no *Pur* sequence was found in the *Y. lipolytica* origin sequences.

Thus, no identifiable short sequence motifs or consensus sequences in *Yarrowia* origins could be detected. This situation can be compared with domains B within *S. cerevisiae ARS*s, which play a crucial role in initiation and can be exchanged between different origins, although they do not display any obvious conserved sequence (Marahrens and Stillman 1992).

Yarrowia Origins Do Not Display Any Characteristic Structural Features

It has been proposed that eukaryotic replication origins are associated with clusters of structural motifs such as SARs, DUEs (Umek and Kowalski, 1988), or bent DNA (Trifonov, 1991), which can be mapped by computer modeling (Eckdahl and Anderson, 1990, Dobbs *et al.*, 1994). As illustrated in Figure 6, A and E, bent DNA is frequently associated with *S. cerevisiae* origins (Snyder *et al.*, 1986; Williams *et al.*, 1988), although it does not seem to be a general feature of the origins in yeast, or may even be dispensable (Marahrens and Stillman, 1992). We display here *ARS*1 and *HMRE* as controls. The curvature of yeast origins was calculated with the wedge model of Trifonov (see MATERIALS AND METHODS) and is displayed as a projection of the spatial path of the molecule for small fragments (Figure 6, A and B) or as a curvature map for longer molecules (ENDS ratio; Figure 6E). When applied to the *Yarrowia* origins (Figure 6, B, D, and E; and our unpublished results), this analysis did not suggest significant DNA bending in the minimal origin fragment.

Regions of low helical stability that facilitate the initial unwinding of the DNA molecule are often found in the vicinity of replication origins (Umek and Kowalski, 1988). We used the thermodynamic library of Breslauer *et al.* (1986) to map the variation of the energy (ΔG) required to unwind the DNA at different origins (Figure 6, C and E). A low ΔG DUE is located close to the ACS of *ARS1* and of the HMRE *ARS*. No such feature is detectable in any of the cloned *Yarrowia* origins (Figure 6, D and E; and our unpublished results), which therefore lack a strong DUE; however, it is noteworthy that easily unwound sequences have been found in the neighboring plasmid sequence that could potentially substitute for DUEs.

Finally, we looked for the presence of putative SARs in *Yarrowia ARS* elements, because they often colocalize with eukaryotic origins (Amati and Gasser, 1988;

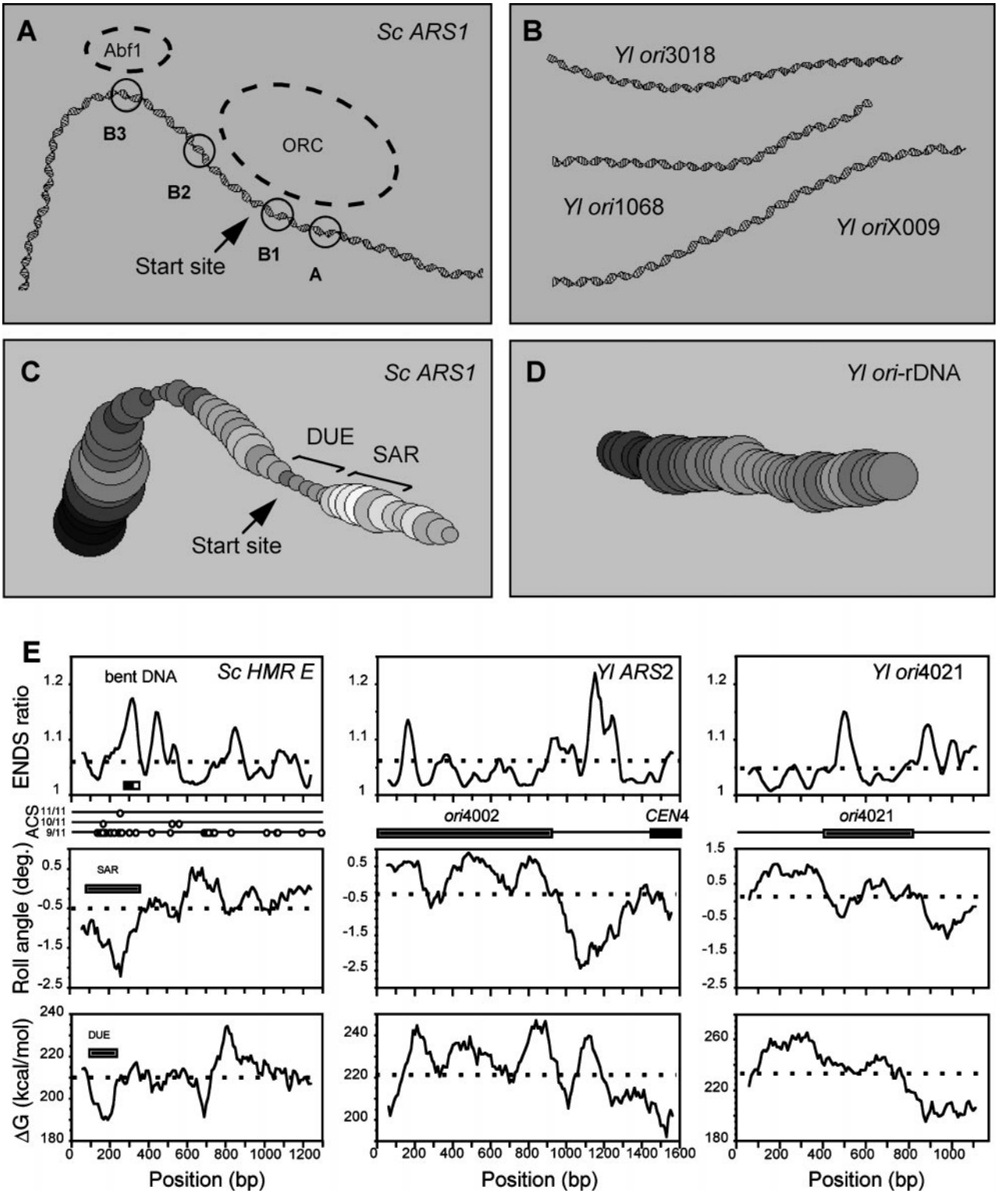


Figure 6. *Y. lipolytica* origins do not contain structural motifs that are frequent in *S. cerevisiae ARS* elements. (A) *S. cerevisiae ARS1* contains a strong, bent DNA region. The 2D projection of the 3D path of the molecule was calculated as described in MATERIALS AND METHODS. The position of the three elements A, B1, B2, and B3 (Marahrens and Stillman, 1992), the binding sites of ORC and ABF1 (Lee and Bell, 1997), and the site of initiation of DNA synthesis (Bielinsky and Gerbi, 1998) are indicated. (B) *Y. lipolytica* origins are not intrinsically curved. The 3D path of the minimal origin region of *ori*3018 (144 bp), *ori*1068 (125 bp), and *ori*X009 (159 bp) was calculated as described for *ARS1.* (C)

Brun *et al.*, 1990; Razin *et al.*, 1991; Du *et al.*, 1995). These sequences are generally AT rich and present a narrow minor groove (Roberge and Gasser, 1992) that can be mapped using the Roll angle, one of the three helical parameters of the double helix (see MATERIALS AND METHODS). As shown in Figure 6, B and E, the region of low Roll angle fits perfectly with the position of the SAR mapped experimentally for *S. cerevisiae ARS* (Amati *et al.*, 1990). When applied to *Y. lipolytica* origins, this analysis did not reveal any significant pattern (Figure 6, D and E). A putative SAR between *ori*4002 and *CEN2* (Figure 6E), as well as a strong bending element (ENDS ratio >1.2) previously detected by gel retardation assay (Matsuoka *et al.*, 1993), were found. Similar "low Roll" regions are predicted between the *ORI* and the *CEN* from the two other chromosomes, suggesting that it may play a role in chromosome maintenance; however this region can be deleted from the plasmids without affecting the transformation frequency (Vernis *et al.*, 1997).

Ectopic Initiation Occurs within Short DNA Fragments

We have shown that short DNA fragments presenting no obvious primary or secondary sequence similarity are able to confer extrachromosomal maintenance to a centromeric plasmid. We examined the question of whether these minimal sequences contain all the information that is necessary and sufficient to function as an origin of replication on the chromosome. We already knew that *ori*1068 is still active when its 3′-flanking region is modified (Vernis *et al.*, 1997). To test whether *Y. lipolytica ORI*s require a particular chromosomal context or additional information to be active in the genome, several origins were moved to another locus and analyzed by 2D gel. Two noncentromeric origins (the 285 bp of *ori*-rDNA and the 158 bp of *ori*X009) and one centromeric origin (the 125-bp minimal *ori*1068) were inserted into the *Sal*I site in pINA214 (pBR322 + *LEU2*) and integrated into the *LEU2* locus in the chromosome after *Bst*XI digestion of the plasmid. A chromosomal *Xho*I–*Sac*II restriction fragment containing the origins was analyzed on 2D gel (Figure 7). A control experiment, performed with the vector alone integrated at the same locus, showed the absence of initiation in this region (Figure 7D), whereas each of the three origins tested generated an initiation signal within this restriction fragment (Figure 7, A–C). Therefore, we conclude that the short *Yarrowia ORI* elements contain the information sufficient to promote initiation at ectopic loci, even when embedded within heterologous (pBR322) sequences.

DISCUSSION

The Y. lipolytica Replicon

It had been shown previously that *ARS* sequences are very rarely obtained from this yeast (Fournier *et al.*, 1991; Matsuoka *et al.*, 1993). We later demonstrated that this low cloning frequency was due to both a centromere and an origin of replication being necessary to establish an *ARS* vector (Fournier *et al.*, 1993; Vernis *et al.*, 1997). Because an origin can be associated with different centromeres to confer extrachromosomal plasmid maintenance, we postulated that any chromosomal origin could therefore be cloned using a centromeric vector (Vernis *et al.*, 1997). The results presented here confirm this hypothesis: several new *ORI* elements from various chromosomal loci were cloned using this strategy. Those tested were shown to be chromosomal initiation sites for DNA replication. Our data suggest that origins of replication are present at every 20–50 kb in the *Y. lipolytica* genome, consistent with the average replicon size of other yeasts (Maundrell *et al.*,1985; Newlon and Theis, 1993).

These origins of replication are well defined genetic elements: the deletion of 240 bp from the *ori*1068 chromosomal locus abolishes initiation, whereas the introduction of bacterial sequences nearby does not (Vernis *et al.*, 1997). We demonstrate here that an element as short as 125–150 bp comprises all the necessary information to initiate DNA replication outside of its natural chromosomal environment. Linker scanning or internal deletion analysis would be required to establish whether all of this sequence is necessary or relevant for *ORI* function

The data presented here indicate that centromeric sequences are not required to initiate DNA replication on the chromosome but are absolutely essential for plasmid maintenance. They may be necessary for any distribution of plasmids at mitosis, which is not the

Figure 6 (facing page). Variation of helical stability and minor groove width along the *S. cerevisiae ARS1* DNA fragment (350 bp). The free energy of the duplex or ΔG (207–227 kcal/mol; circles of increasing size) and the mean Roll angle (−1 to −0.1°; gray levels of increasing intensity) were calculated as described in MATERIALS AND METHODS for a 120-bp window sliding along the molecule. The position of the SAR mapped by Amati *et al.*, 1988 (low Roll angle) and of the DUE (low ΔG) are indicated. (D) Spatial path and helical parameters of *Y. lipolytica* minimal *ori*-rDNA fragment (284 bp). The mean Roll angle varies from −0.3 to 0.5°, and the ΔG ranges from 215 to 226 kcal/mol. (E) Structural maps of large fragments containing *S. cerevisiae* HMRE *ARS*, *Y. lipolytica ARS*2, and *Y. lipolytica ori*4021. Like *ARS1*, the ACS of the HMRE *ARS* is flanked by a bent motif bearing binding sites for the auxiliary factors RAP1 (dark box) and ABF1 (open box) and by a DUE. The SAR fragment (Amati *et al.*, 1988) maps perfectly to the region of low Roll angle. In all the graphs, the dotted lines correspond to the average value. *Ori*4021 is shown in its pBR322 environment, and the other two origins are at their chromosomal locus.

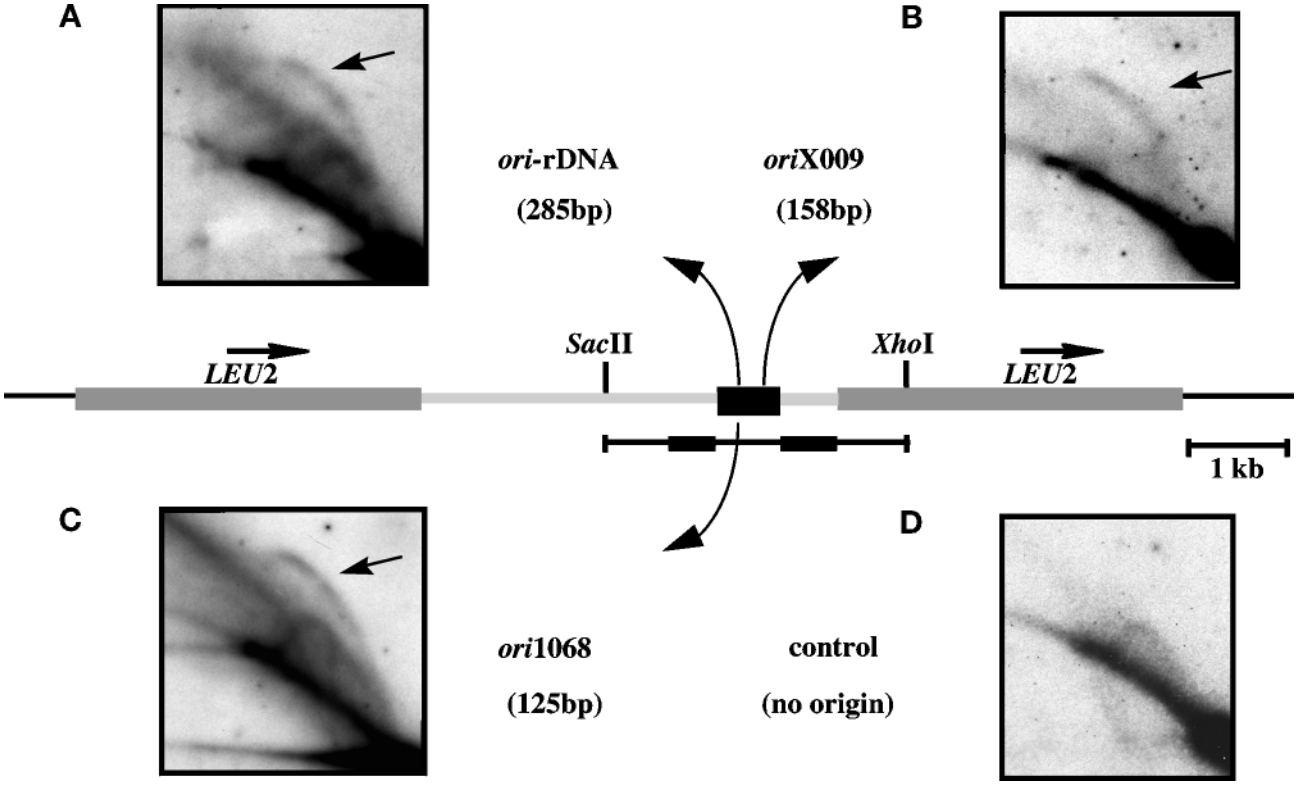


Figure 7. Ectopic initiation occurs within short DNA fragments. (A–C) Three different origins (A, *ori*-rDNA; B, *ori*X009; C, *ori*1068) were ligated into pINA214; the resulting plasmids were digested with *Bst*XI and integrated into the *LEU2* locus. The resulting genomic restriction map is shown, with the position of the sites used to digest the DNA for 2D gel analysis. The gels were blotted and hybridized with a pBR322 probe, indicated as black boxes. (D) The same experiment was performed with pINA214 as a control: hybridization of the 2D gel showed the absence of initiation at the *LEU2* locus. Notice that the diagonal signal visible on these images is present on most chromosomal 2D gels performed in *Yarrowia* and may correspond to degradation products.

case in *S. cerevisiae*. Alternatively, they may function as helper sequences for the replication origins, perhaps through a modification of their chromatin structure or of their subnuclear localization. These two possibilities are not mutually exclusive, but several lines of evidence are in favor of the second interpretation. Indeed, it has been reported that *S. cerevisiae ARS* elements bind the nuclear scaffold (Amati *et al.*, 1990). These origins contain AT-rich stretches with a narrow minor grove (Figure 6) that could target the plasmids to the replication foci within the nucleus (Cook, 1991; Pasero *et al.*, 1997). As shown by computer modeling (Figure 6) or by the χ^2 test (our unpublished results), *Y. lipolytica* origins do not contain any similar structures, but the centromeric regions do. Therefore, the putative role of centromeres on a plasmid may be to direct plasmids to the replication centers. The SAR activity of *Y. lipolytica ORI* and *CEN* fragments still has to be tested to confirm this hypothesis, as well as the replacement of the *CEN* on the plasmid with an exogenous SAR fragment (Vernis, unpublished observations), but this major difference between *Y. lipolytica* and *S. cerevisiae ARS* will certainly prove very useful for investigating the putative role of SARs in the definition of eukaryotic origins of replication (Hyrien *et al.*, 1997).

Y. lipolytica Origins Do Not Share Any Essential Consensus Motif

In the initially cloned 1.3-kb *ARS*18 and 2.3-kb *ARS*68 fragments, we observed several 9/11 matches to the *S. cerevisiae* ACS, which may be why they were able to replicate in the budding yeast (Fournier *et al.*, 1993); however, these ACS are not present within the minimal *ORI* sequences. We then identified a short 8-bp stretch present in all but two sequenced origins; however, three observations suggest that this sequence is not essential for origin function. 1) Mutation of this sequence does not affect transformation frequency; 2) two origins (*ori*X009 and *ori*X096), which do not harbor a perfect match to this consensus, allow extrachromosomal replication on a centromeric plasmid; and 3) one of these two origins (*ori*X009) is still able to initiate DNA replication after being moved to an ectopic chromosomal location, where it is flanked by bacterial

plasmid sequences (Figure 7B). Most of the origins analyzed harbor several degenerated copies of this consensus, however, and in *S. cerevisiae* the replicative property is conserved when several partial ACS are present within the *ARS* (Zweifel and Fangman, 1990). We examined all exact and degenerated copies of the consensus in the *Yarrowia* origins (32 occurrences) and found a more degenerate motif: WDMRWNYH (R = purine; Y = pyrimidine; M = A or C; W = A or T; D = A or G or T). Because some of the analyzed sequences share only five bases, it seems very unlikely that this motif represents a biologically significant consensus. Theis and Newlon (1997) found that ORC can bind a 9/11 match to the *S. cerevisiae* ACS, and this lead them to redefine the consensus on a broader length (17 bp instead of 11 bp). We therefore searched the flanking sequences of all exact matches to the *Y. lipolytica* putative consensus and did not find any other conserved bases that could allow a broader definition of this element.

We devised a program to look at the distribution of bases within the sequences of the origins, and no bias was found. We also analyzed the 2675-bp *ura4* locus of *S. pombe* (67.6% AT), which harbors the two initiation sites *ars3002* and *ars3003* (Dubey *et al.*, 1994). The result is significant (χ^2 of 23.76 vs. 21.03 in the table). Surprisingly, when this analysis is performed on each *ARS* separately (821 bp of *ars3002* and 543 bp of *ars3003*), no bias is observed. A similar observation was made for the *Y. lipolytica ori*3018 environment (see above). In this case, the origin efficiency should be very dependent on the nature of the flanking sequences, as postulated for some metazoan origins (Larner *et al.*, 1997) and should be very sensitive to chromosomal position effects. In *S. pombe* the origin efficiency is indeed affected by the presence of exogenous sequences, and plasmid maintenance is sometimes associated with a multimerization of the origin in vivo (Zhu *et al.*, 1994). In *Y. lipolytica*, we never observed such an alteration of plasmid structure, which suggests that all the information necessary and sufficient for initiation is found within the origin. This is consistent with the observation that such a sequence can be moved to the *LEU2* locus and retain origin activity; however, we did not compare the initiation efficiencies at the natural and at this ectopic locus and therefore cannot completely rule out a possible effect of the genomic environment on origin activity. It seems at least likely that a DUE, if absent from the cloned origins, should be present in the vicinity of the origins on the chromosome to facilitate the entry of the replication machinery (Lin and Kowalski, 1997).

A situation similar to that of *Y. lipolytica* origins has been found in *Physarum*, where two origins have been described in the actin genes promoters (Bénard *et al.*, 1996). In both loci—962-bp *actB* with 53.3% AT, and 1049-bp *actC* with 56.8% AT—no bias was observed in the distribution of AT-rich sequences. In both organisms the AT content of origins is quite low, the bases seem to be randomly distributed, and no canonical origin consensus nor sequences similarity between the origins are found. It could be interesting to check whether *Y. lipolytica* origins are within or near promoter regions.

Yarrowia As a Model to Analyze ORC–Origin Interaction

As mentioned above, the ORC is highly conserved among eukaryotes. It is required for the initiation of DNA replication in *S. cerevisiae*, *S. pombe*, *Xenopus*, or *Drosophila* (for review, see Diffley, 1996), although the replication origins are structurally very different in these organisms. Several explanations of this apparent paradox have been proposed (Burhans and Huberman, 1994; Diffley, 1996; Larner *et al.*, 1997). One tempting idea is that any DNA sequence that is able to bind ORC and that is located in a chromatin environment permissive for initiation becomes an origin. In this case, the higher-order organization of chromatin would be the main determinant of origin function. Chromosome architecture was indeed shown to dictate site-specific initiation in *Xenopus* egg extracts (Lawlis *et al.*, 1996; for review, see Hyrien *et al.*, 1997), and the random initiation observed in the rDNA of early *Xenopus* embryos is subsequently restricted to nontranscribed regions after activation of transcription (Hyrien *et al.*, 1995). Alternatively, the binding of ORC could be strictly sequence-specific in somatic cells but could also occur at degenerated sites when the ratio of ORC to DNA is very high, as in early embryos. It is therefore essential to characterize the sequences that are recognized by this complex in systems other than *S. cerevisiae*. This is technically very difficult in higher eukaryotes, where the initiation zones are large regions. Interestingly, Sanchez *et al.* (1998b) showed that several proteins bind to *ARS* elements in *S. pombe*; however, the two proteins identified so far do not correspond to the initiator protein. Similar work in *Y. lipolytica* would be valuable, because origins correspond to much shorter and well defined genetic elements. The identification at *Y. lipolytica* origins of a footprint that changes during the cell cycle would certainly be an important step toward the characterization of the sequences recognized by ORC on eukaryotic chromosomes. This is a further illustration that unicellular eukaryotic organisms different from *S. cerevisiae* can provide useful models for exploring mechanisms at replication origins.

ACKNOWLEDGMENTS

We thank P. Trécourt for the development of the statistical program used in this article. This work was supported by grants from Centre National de la Recherche Scientifique (URA547 and URA1189) and Institut National de de la Recherche Agronomique (UR51216). L.V. was supported by a fellowship from the French Ministère de la

Recherche; P.P. was the recipient of a fellowship from the Fondation pour la Recherche Médicale and a European Molecular Biology Organization long-term fellowship.

REFERENCES

Amati, B.B., and Gasser, S.M. (1988). Chromosomal *ARS* and *CEN* elements bind specifically to the yeast nuclear scaffold. Cell *54*, 967–978.

Amati, B., Pick, L., Laroche, T., and Gasser, S.M. (1990). Nuclear scaffold attachment stimulates, but is not essential for *ARS* activity in *Saccharomyces cerevisiae*: analysis of the *Drosophila ftz* SAR. EMBO J. *9*, 4007–4016.

Barth, G., and Gaillardin, C. (1996). *Yarrowia lipolytica*. In: Nonconventional Yeasts in Biotechnology: A Handbook, ed. K. Wolf, Springer-Verlag: Heidelberg, 313–388.

Bénard, M., Lagnel, C., Pallotta, D., and 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.

Bénard, M., Lagnel, C., and Pierron, G. (1995). Site-specific initiation of DNA replication within the nontranscribed spacer of *Physarum* rDNA. Nucleic Acids Res. *23*, 1447–1453.

Bergemann, A.D., and Johnson, E.M. (1992). The HeLa Pur factor binds single-stranded DNA at a specific element conserved in gene flanking regions and origins of DNA replication. Mol. Cell. Biol. *12*, 1257–1265.

Bielinsky, A.K., and Gerbi, S.A. (1998). Discrete Start Sites for DNA Synthesis in the Yeast *ARS*1 Origin. Science *279*, 95–98.

Bolshoy, A., McNamara, P., Harrington, R.E., and Trifonov, E.N. (1991). Curved DNA without A-A: experimental estimation of all 16 DNA wedge angles. Proc. Natl. Acad. Sci. USA *88*, 2312–2316.

Breslauer, K.J., Franck, R., Blöcker, H., and Marky, L.A. (1986). Predicting DNA duplex stability from the base sequence. Proc. Natl. Acad. Sci. USA *83*, 3746–3750.

Brun, C., Dang, Q., and Miassod, R. (1990). Studies of an 800-kb DNA stretch of the *Drosophila* X chromosome: comapping of a subclass of scaffold-attached regions with sequences able to replicate autonomously in *Saccharomyces cerevisiae*. Mol. Cell. Biol. *10*, 5455–5463.

Burhans, W.C., and Huberman, J.A. (1994). DNA replication origins in animal cells: a question of context? Science *263*, 639–640.

Burhans, W.S., Vassilev, L.T., Caddle, M.S., Heintz, N.H., DePamphilis, M.L. (1990). Identification of an origin of bidirectional DNA replication in mammalian chromosomes. Cell *62*, 955–965.

Caddle, M.S., and Calos, M.P. (1994). Specific initiation at an origin of replication from *Schizosaccharomyces pombe*. Mol. Cell. Biol. *14*, 1796–1805.

Casarégola, S., Feynerol, C., Diez, M., Fournier, P., and Gaillardin, C. (1997). Genomic organization of the yeast *Yarrowia lipolytica*. Chromosoma *106*, 380–390.

Cook, P.R. (1991). The nucleoskeleton and the topology of replication. Cell *66*, 627–635.

DePamphilis, M. L. (1996). Origins of DNA replication. In: DNA Replication in Eukaryotic Cells, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 45–86.

Dhar, S.K., Chardhury, N.R., Mittal, V., Bhattacharya, A., and Bhattacharya, S. (1996). Replication initiates at multiple dispersed sites in the ribosomal DNA plasmid of the protozoan parasite *Entamoeba histolytica*. Mol. Cell. Biol. *16*, 2314–2324.

Diffley, J.F., Cocker, J.H., Dowell, S.J., and Rowley, A. (1994). Two steps in the assembly of complexes at yeast replication origins in vivo. Cell *78*, 303–316.

Diffley, J.F., Cocker, J.H., Dowell, S.J., Harwood, J., and Rowley, A. (1995). Stepwise assembly of initiation complexes at budding yeast replication origins during the cell cycle. J. Cell Sci. Suppl. *19*, 67–72.

Diffley, J.F.X. (1996). Once and only once upon a time: specifying and regulating origins of DNA replication in eukaryotic cells. Genes Dev. *10*, 2819–2830.

Dijkwel, P.A., Vaughn, J.P., and Hamlin, J.L. (1994). Replication initiation sites are distributed widely in the amplified CHO dihydrofolate reductase domain. Nucleic Acids Res. *22*, 4989–4996.

Dobbs, D.L., Shaiu, W.L., and Benbow, R.M. (1994). Modular sequence elements associated with origin regions in eukaryotic chromosomal DNA. Nucleic Acids Res. *22*, 2479–2489.

Du, C., Sanzgiri, R.P., Shaiu, W-L., Choi, J-K., Hou, Z., Benbow, R.M., and Dobbs, D.L. (1995). Modular structural elements in the replication origin region of *Tetrahymena* rDNA. Nucleic Acids Res. *23*, 1766–1774.

Dubey, D.D., Kim, S.M., Todorov, I.T., and Huberman, J.A. (1996). Large, complex modular structure of a fission yeast DNA replication origin. Curr. Biol. *6*, 467–473.

Dubey, D.D., Zhu, J.G., Carlson, D.L., Sharma, K., and Huberman, J.A. (1994). Three *ARS* elements contribute to the *ura4* replication origin region in the fission yeast, *Schizosaccharomyces pombe*. EMBO J. *13*, 3638–3647.

Eckdahl, T.T., and Anderson, J.N. (1987). Computer modelling of DNA structures involved in chromosome maintenance. Nucleic Acids Res. *15*, 8530–8545.

Eckdahl, T.T., and Anderson, J.N. (1990). Conserved DNA structures in origins of replication. Nucleic Acids Res. *18*, 1609–1612.

Fabiani, L., Frontali, L., and Newlon, C.S. (1996). Identification of an essential core element and stimulatory sequences in a *Kluyveromyces lactis ARS* element, *KARS101*. Mol. Microbiol. *19*, 757–766.

Fournier, P., Abbas, A., Chasles, M., Kudla, B., Ogrydziak, D.M., Yaver, D., Xuan, J-W., Peito, A., Ribet, A-M., Feynerol, C., He, F., and Gaillardin, C. (1993). Colocalization of centromeric and replicative functions on autonomously replicating sequences isolated from the yeast *Yarrowia lipolytica*. Proc. Natl. Acad. Sci. USA *90*, 4912–4916.

Fournier, P., Gaillardin, C., Persuy, M.A., Klootwijk, J., and van Heerikhuizen, H. (1986). Heterogeneity in the ribosomal family of the yeast *Yarrowia lipolytica*: genomic organization and segregation studies. Gene *42*, 273–282.

Fournier, P., Guyaneux, L., Chasles, M., and Gaillardin, C. (1991). Scarcity of *ARS* sequences isolated in a morphogenesis mutant of the yeast *Yarrowia lipolytica*. Yeast *7*, 25–36.

Huberman, J.A. (1994). Analysis of DNA replication origins and directions by two-dimensional gel electrophoresis. In: The Cell Cycle: A Practical Approach, ed. P. Fantes and R.F. Brooks, Oxford, UK: Oxford University Press, 213–234.

Hyrien, O., Maric, C., and Lucas, I. (1997). Role of nuclear architecture in the initiation of eukaryotic DNA replication. Biochimie *79*, 541–548.

Hyrien, O., Maric, C., and Méchali, M. (1995). Transition in specification of embryonic metazoan DNA replication origins. Science *270*, 994–997.

Jallepalli, P.V., and Kelly, T.J. (1997). Cyclin-dependent kinase and initiation at eukaryotic origins: a replication switch? Curr. Opin. Cell Biol. *9*, 358–363.

Larner, J.M., Lee, H., and Hamlin, J.L. (1997). S phase damage sensing checkpoints in mammalian cells. Cancer Surv. *29*, 25–45.

Lawlis, S.J., Keezer, S.M., Wu, J-R., and Gilbert, D.M. (1996). Chromosome architecture can dictate site-specific initiation of DNA replication in *Xenopus* egg extracts. J. Cell Biol. *135*, 1207–1218.

Lee, D.G., and Bell, S.P. (1997). Architecture of the yeast origin recognition complex bound to origins of DNA replication. Mol. Cell. Biol. *17*, 7159–7168.

Lin, S., and Kowalski, D. (1997). Functional equivalency and diversity of *cis*-acting elements among yeast replication origins. Mol. Cell. Biol. *17*, 5473–5484.

Linskens, M.H., and Huberman, J.A. (1988). Organization of replication of ribosomal DNA in *Saccharomyces cerevisiae*. Mol. Cell. Biol. *8*, 4927–4935.

Marahrens, Y., and Stillman, B. (1992). A yeast chromosomal origin of DNA replication defined by multiple functional elements. Science *255*, 817–823.

Marilley, M., and Pasero, P. (1996). Common DNA structural features exhibited by eukaryotic ribosomal gene promoters. Nucleic Acids Res. *24*, 2204–2211.

Matsuoka, M., Matsubara, M., Daidoh, H., Imanaka, T., Uchida, K., and Aiba, S. (1993). Analysis of regions essential for the function of chromosomal replicator sequences from *Yarrowia lipolytica*. Mol. Gen. Genet. *237*, 327–333.

Maundrell, K., Hutchinson, A., and Shall, S. (1988). Sequence analysis of *ARS* elements in fission yeast. EMBO J. *7*, 2203–2209.

Maundrell, K., Wright, A.P.H., Piper, M., and Shall, S. (1985). Evaluation of heterologous *ARS* activity in *S. cerevisiae* using cloned DNA from *S. pombe*. Nucleic Acids Res. *13*, 3711–3722.

Natale, D.A., Umek, R.M., and Kowalski, D. (1993). Ease of DNA unwinding is a conserved property of yeast replication origins. Nucleic Acids Res. *21*, 555–560.

Newlon, C.S., Collins, I., Dershowitz, A., Deshpande, A.M., Greenfeder, S.A., Ong, L.Y., and Theis, J.F. (1993). Analysis of replication origin function on chromosome III of *Saccharomyces cerevisiae*. Cold Spring Harbor Symp. Quant. Biol. *58*, 415–423.

Newlon, C.S., and Theis, J.F. (1993). The structure and function of yeast *ARS* elements. Curr. Opin. Genet. Dev. *3*, 752–758.

Pasero, P., Braguglia, D., and Gasser, S.M. (1997). ORC-dependent and origin-specific initiation of DNA replication at defined foci in isolated yeast nuclei. Genes Dev. *11*, 1504–1518.

Pasero, P., Sjakste, N., Blettry, C., Got, C., and Marilley, M. (1993). Long-range organization and sequence-directed curvature of *Xenopus laevis* satellite 1 DNA. Nucleic Acids Res. *21*, 4703–4710.

Razin, S.V., Vassetzky, Y.S., and Hancock, R. (1991). Nuclear matrix attachment regions and topoisomerase II binding and reaction sites in the vicinity of a chicken DNA replication origin. Biochem. Biophys. Res. Commun. *177*, 265–270.

Roberge, M., and Gasser, S.M. (1992). DNA loops: structural and functional properties of scaffold-attached regions. Mol. Microbiol. *6*, 419–423.

Rowles, A, and Blow, J. (1997). Chromatin proteins involved in the initiation of DNA replication. Curr. Opin. Genet. Dev. *7*, 152–157.

Rowley, A., Cocker, J.H., Harwood, J., and Diffley, J.F. (1995). Initiation complex assembly at budding yeast replication origins begins with the recognition of a bipartite sequence by limiting amounts of the initiator, ORC. EMBO J. *14*, 2631–2641.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Sanchez, J.A., Kim, S-M., and Huberman, J.A. (1998a). Ribosomal DNA replication in the fission yeast, *Schizosaccharomyces pombe*. Exp. Cell Res. *238*, 220–230.

Sanchez, J.P., Murakami, Y., Huberman, J.A., and Hurwitz, J. (1998b). Isolation, characterization, and molecular cloning of a protein (Abp2) that binds to a *Schizosaccharomyces pombe* origin of replication (*ars3002*). Mol. Cell. Biol. *18*, 1670–1681.

Schmid-Berger, N., Schmid, B., and Barth, G. (1994). Ylt1, a highly repetitive retrotransposon in the genome of the dimorphic fungus *Yarrowia lipolytica*. J. Bacteriol. *176*, 2477–2482.

Shinomiya, T., and Ina, S. (1994). Mapping an initiation region of DNA replication at a single-copy chromosomal locus in *Drosophila melanogaster* cells by two-dimensional gel methods and PCR-mediated nascent-strand analysis: multiple replication origins in a broad zone. Mol. Cell. Biol. *14*, 7394–7403.

Snyder, M., Buchman, A.R., and Davis, R.W. (1986). Bent DNA at a yeast autonomously replicating sequence. Nature *324*, 87–89.

Tanaka, T., Knapp, D., and Nasmyth, K. (1997). Loading of an Mcm protein onto DNA replication origins is regulated by Cdc6p and CDKs. Cell *90*, 649–660.

Theis, J.F., and Newlon, C.S. (1997). The *ARS309* chromosomal replicator of *Saccharomyces cerevisiae* depends on an exceptional *ARS* consensus sequence. Proc. Natl. Acad. Sci. USA *94*, 10786–10791.

Trifonov, E.N. (1991). DNA in profile. Trends Biochem. Sci. *16*, 467–470.

Umek, R.M., and Kowalski, D. (1988). The ease of DNA unwinding as a determinant of initiation at yeast replication origins. Cell *52*, 559–567.

van Heerikhuizen, H., Ykema, A., Klootwijk, J., Gaillardin, C., Ballas, C., and Fournier, P. (1985). Heterogeneity in the rRNA genes of the yeast *Yarrowia lipolytica*: cloning and analysis of two size classes of repeats. Gene *39*, 213–222.

Vernis, L., Abbas, A., Chasles, M., Gaillardin, C.M., Brun, C., Huberman, J.A., and Fournier, P. (1997). An origin and a centromere are both needed to establish a replicative plasmid in the yeast *Yarrowia lipolytica*. Mol. Cell. Biol. *17*, 1995–2004.

Vita-Pearlman, V.J., Gunaratne, P.H., and Chinault, A.C. (1993). Analysis of a replication initiation sequence from the adenosine deaminase region of the mouse genome. Mol. Cell. Biol. *13*, 5931–5942.

Wesolowski, M., Algeri, A., and Fukuhara, H. (1981). Gene organization of the mitochondrial DNA of yeasts: *Kluyveromyces lactis* and *Saccharromycopsis lipolytica*. Curr. Genet. *3* 157–162.

Williams, J.S., Eckdahl, T.T., and Anderson, J.N. (1988). Bent DNA functions as a replication enhancer in *Saccharomyces cerevisiae*. Mol. Cell. Biol. *8*, 2763–2769.

Wohlgemuth, J.G., Bulboaca, G.H., Moghadam, M., Caddle, M.S., and Calos, M.P. (1994). Physical mapping of origins of replication in the fission yeast *Schizosaccharomyces pombe*. Mol. Biol. Cell *5*, 839–849.

Xuan, J.W., Fournier, P., and Gaillardin, C. (1988). Cloning of the *LYS5* gene encoding saccharopine dehydrogenase from the yeast *Yarrowia lipolytica* by targeted integration. Curr. Genet. *14*, 19–21.

Zhu, J.G., Carlson, D.L., Dubey, D.D., Sharma, K., and Huberman, J.A. (1994). Comparison of the two major *ARS* elements of the *ura4* replication origin region with other *ARS* elements in the fission yeast, *Schizosaccharomyces pombe*. Chromosoma *103*, 414–422.

Zimmermann, M., and Fournier, P. (1996). Electrophoretic karyotyping of yeasts. In: Nonconventional Yeasts in Biotechnology: A Handbook, ed. K. Wolf, Springer-Verlag: Heidelberg, 101–116.

Zweifel, S.G., and Fangman, W.L. (1990). Creation of *ARS* activity in yeast through iteration of nonfunctional sequences. Yeast *6*, 179–186.