Deletion Mutations Affecting Autonomously Replicating Sequence ARS1 of Saccharomyces cerevisiae

S. E. CELNIKER,† K. SWEDER, F. SRIENC, J. E. BAILEY, AND J. L. CAMPBELL*
Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125

Received 4 June 1984/Accepted 7 August 1984

DNAs that contain specific yeast chromosomal sequences called ARSs transform Saccharomyces cerevisiae at high frequency and can replicate extrachromosomally as plasmids when introduced into S. cerevisiae by transformation. To determine the boundaries of the minimal sequences required for autonomous replication in S. cerevisiae, we have carried out in vitro mutagenesis of the first chromosomal ARS described, ARS1. Rather than identifying a distinct and continuous segment that mediates the ARS+ phenotype, we find three different functional domains within ARS1. We define domain A as the 11-base-pair (bp) sequence that is also found at most other ARS regions. It is necessary but not sufficient for high-frequency transformation. Domain B, which cannot mediate high-frequency transformation, or replicate by itself, is required for efficient, stable replication of plasmids containing domain A. Domain B, as we define it, is continuous with domain A in ARS1, but insertions of 4 bp between the two do not affect replication. The extent of domain B has an upper limit of 109 bp and a lower limit of 46 bp in size. There is no obvious sequence homology between domain B of ARS1 and any other ARS sequence. Finally, domain C is defined on the basis of our deletions as at least 200 bp flanking domain A on the opposite side from domain B and is also required for the stability of domain A in S. cerevisiae. The effect of deletions of domain C can be observed only in the absence of domain B, at least by the assays used in the current study, and the significance of this finding is discussed.

As in higher organisms, chromosome replication in Saccharomyces cerevisiae initiates at multiple sites and is bidirectional. The DNA is replicated only during the S phase, and activation of individual replicons is thought to occur according to a specific temporal program, with each replicon activated only once per cell cycle (for review, see reference 3). There is no direct evidence in any eucaryote that chromosomal replication initiates at the same chromosomal sequence during each cell cycle. However, indirect evidence is available in S. cerevisiae. Yeast sequences have been isolated that confer on any colinear DNA the ability to be maintained in S. cerevisiae as autonomously replicating, albeit highly unstable, plasmids (20, 41, 44). The sequences responsible for autonomous replication are designated ARS. Estimates of the number of ARS elements in the S. cerevisiae genome (1, 7) agree with the number of initiation sites for DNA replication estimated by fiber autoradiography and electron microscopy (reviewed in reference 3). ARS elements have been classified by nucleic acid hybridization as either unique or repetitive (6, 8, 41, 45). In addition to S. cerevisiae DNA, DNAs from a wide variety of eucaryotes including Neurospora crassa, Dictyostelium discoideum, Caenhorabditis elegans, Drosophila melanogaster, Zea mays (42), humans (31), Xenopus mitochondrial DNA (51), and Tetrahymena rDNA (25) also promote high-frequency transformation of recombinant DNA plasmids in yeasts. Thus it is even possible that initiation of DNA replication occurs at sequences which are similar in many eucaryotes. But are such ARS elements origins of replication?

A eucaryotic chromosomal origin of replication should contain sequences that allow the initiation of DNA synthesis, initiation of bidirectional replication, and the temporal regulation of initiation within the S phase. We have shown, by in vitro replication, that ARS plasmids contain the necessary information to direct the initiation of DNA synthesis at a unique site (5) and that subsequent synthesis is sequential and bidirectional. Fangman et al. (11) have recently shown that the yeast ARS plasmids replicate only once during each S phase and also contain the information necessary to determine the time of activation of replication within the S phase. These sequences also replicate under the control of the CDC genes that are required for chromosomal replication (52). The only substantive evidence against ARS elements being origins of replication comes from the fact that specific sequences do not seem to be required in Xenopus embryos for cell-cycle-regulated DNA replication (e.g., see reference 17).

Hoping to gain a better understanding of the function of ARSs, we have investigated the organization of the S. cerevisiae ARS1 element, first identified by Struhl et al. (44). Stinchcomb et al. (40, 41) located ARS1 to an 838-base-pair (bp) EcoRI-HindIII fragment. Further studies showed that a smaller fragment of 600 bp was not sufficient for ARS function (40, 48). We have extended these studies by systematically generating mutations in vitro in the cloned ARS1 element and then introducing these mutant DNAs into S. cerevisiae cells by transformation. We have assayed ARS activity by the ability of the plasmid to transform the S. cerevisiae cells at high frequency and to be maintained as extrachromosomal DNA. Instead of finding, as expected by analogy to oriC of Escherichia coli, that ARS activity can be described in terms of a single continuous minimal DNA segment, we have found that ARS1 consists of different domains, each of which affects only part of the ARS+ phenotype. We also describe a new method for determining the plasmid copy number.

MATERIALS AND METHODS

Materials. Acrylamide, bisacrylamide, and bisacryloyl-cystamine were obtained from Bio-Rad Laboratories; Sea-
Kem agarose and low-melting-temperature agarose (Sea Plaque) were obtained from FMC Corp., Marine Colloids Div.; [α-32P]dNTPs (400 to 3,000 Ci/mmol) were obtained from Amersham Corp.; [γ-32P]ATP (8,000 Ci/mmol) was obtained from ICN Pharmaceuticals Inc.; restriction enzymes were obtained from New England Biolabs; DNA modifying enzymes were obtained from Boehringer Mannheim Biochemicals, Bethesda Research Laboratories, and Sigma Chemical Co.; and T4 poly nucleotide kinase and T4 DNA ligase were gifts from C. C. Richardson, Harvard University, Boston, Mass. EcoRI and HindIII linkers were from Collaborative Research, Inc.

Strains, plasmids, and media. The S. cerevisiae strain used for yeast transformation, SS111 (a trpl-289 ura3-1 ade2-10 his3-523 ade2-10 gal12 ino can5) was a gift from S. Scherer, California Institute of Technology, Pasadena. The yeast strain used to study the plasmid genes (see Fig. 5) was YM603 (a ura3-52 lys2-801 met his3 ade2-10 regl-501), which was a gift from M. Johnston, Washington University, St. Louis, Mo., and which contains a mutation (regl-501) that renders the strain insensitive to catabolite repression by glucose. Strain D603 is a homozygous a/a diploid prepared from strain YM603. Plasmid pLGS5 is described by Guarante et al. (15). All plasmids used in these studies are described below (see Fig. 1 and 6).

Complete medium contained 1% yeast extract, 2% peptone, and 2% glucose (YPD medium). Minimal medium was 0.64% yeast nitrogen base–2% glucose–0.002% adenine sulfate–0.002% uracil and was supplemented with purified amino acids (SD medium).

E. coli MC1061 (4) was grown in either Vogel-Bonner medium (50) or L broth (28). For selective growth, ampicillin was added to 35 to 50 μg/ml.

DNA preparations. Yeast plasmid DNA was prepared by the method of Nasmyth and Reed (34). To identify yeast plasmids, a sample of the DNA was electrophoresed on a 0.8% agarose gel as described below, transferred to nitrocellulose, and then localized by hybridization with pBR322 [32P]DNA (32, 37).

Plasmid DNA was prepared on a large scale (1 to 4 liters) from stationary phase E. coli cells grown in Vogel-Bonner medium (10).

Yeast transformations. All yeast transformation efficiencies were determined with 4 μg of plasmid DNA purified as described in reference 10. Transformation was by a modification of the method of Ito et al. (21) as described by Kuo and Campbell (26).

Gel electrophoresis. DNA was electrophoresed in 0.8 to 3% agarose cells containing TAE buffer (40 mM Tris-hydrochloride [pH 7.5], 10 mM sodium acetate, 2 mM EDTA) at 30 to 60 V for 12 to 16 h. The DNA was visualized by staining with ethidium bromide.

Polyacrylamide gels were used to resolve DNA fragments less than 600 bp in size. The acrylamide–N,N′-methylene bisacrylamide ratio was 29:1. and the concentration of acrylamide varied from 8 to 25%. Gels were prepared, and electrophoresis was carried out in Tris-borate buffer (TBE: 88 mM Tris-hydrochloride, 89 mM boric acid, 2.5 mM Na2-EDTA [pH 8.3]). Some DNA fragments for sequencing were isolated from acrylamide gels with the cleavable cross-linker bisacylamidostain in the ratio 19:1, acrylamide to cross-linker (16).

Construction of deletion and insertion mutants. DNA molecules were linearized with the appropriate restriction enzyme and extracted with phenol and then chloroform-isoamyl alcohol (24:1); the DNA was ethanol precipitated. Linear DNA (1 μg) was incubated for 30 min at 37°C in a reaction mix (0.3 ml) containing 280 mM NaCl, 30 mM sodium acetate (pH 4.4), 4.5 mM ZnCl2, and 30 to 67 U of SI nuclease (12). The reaction was terminated by the addition of sodium acetate to a final concentration of 0.3 M. The DNA was precipitated with EtOH and resuspended in a 200-μl ligation reaction mixture that contained 50 mM Tris-hydrochloride (pH 7.5), 10 mM MgCl2, 1 mM dithiothreitol, 0.1 mM ATP, and 1 μl of ligase. The reaction was incubated for 16 h at 12°C.

DNA molecules were linearized and prepared as described for the SI nuclease digestion. DNA (1 μg) was incubated for 30 min at 25°C in a reaction mix (25 μl) containing 0.5 M Tris-hydrochloride (pH 7.8), 5 mM MgCl2, 5 mM dithiothreitol, 50 μg of bovine serum albumin per ml, 0.1 mM each dATP, dTTP, dCTP, and dGTP, and 1 U of the Klenow fragment of E. coli DNA polymerase I. The reaction was terminated, and the DNA was prepared and ligated as described above.

Construction of plasmids containing Bal31-generated deletions. Plasmid DNA (10 μg) was cut with a 10-fold excess of restriction enzyme at 37°C for 2 h in an 0.05-M reaction mixture containing 60 mM NaCl, 7 mM Tris-hydrochloride (pH 7.4), 7 mM MgCl2, and 150 μg of bovine serum albumin per ml. The restriction enzyme was inactivated by incubation at 70°C for 10 min. The reaction mix was then made 12 mM CaCl2, 12 mM MgCl2, 20 mM Tris-hydrochloride (pH 8.0), 1 mM EDTA, and 0.2 M NaCl, and the final volume was brought to 0.1 ml. The appropriate concentration of the double-strand exonuclease Pseudomonas Bal31 was added (27). At 1-min intervals, 10-μl samples were removed and the reaction was stopped by the addition of EGTA [ethyleneglycol-bis(β-aminooethyl ether)-N,N′-tetraacetic acid] to a final concentration of 0.02 M. A 20% portion of the sample was digested with restriction enzymes and analyzed by gel electrophoresis. The remainder was stored at −20°C in EtOH.

The Bal31-digested DNA was collected by centrifugation and resuspended in 0.05 ml of 0.05 M Tris-hydrochloride (pH 9.0). The DNA was dephosphorylated by the addition of 0.01 U of calf intestinal alkaline phosphatase, followed by incubation for 30 min at 37°C. The phosphatase was inactivated by the addition of 1 μl of 0.2 M sodium nitrilotriacetate for 10 min at 37°C, and DNA was precipitated with ethanol. Oligonucleotide linkers (EcoRI or HindIII) were phosphorylated in a reaction mixture containing 50 mM Tris-hydrochloride (pH 9.0), 10 mM MgCl2, 0.1 mM EDTA, 0.1 mM spermidine, 1 mM dithiothreitol, 0.05 mM ATP; and 1 μl of T4 polynucleotide kinase for 30 min at 37°C. The reaction was stopped by freezing at −20°C.

Phosphorylated linker (100 ng) was then ligated to the Bal31-digested, dephosphorylated DNA in a 0.25-ml reaction mix which contained 50 mM Tris-hydrochloride (pH 7.5), 10 mM MgCl2, 1 mM dithiothreitol, and 0.1 mM ATP for 3 h at 12°C. The ligase was inactivated by heating for 10 min at 70°C. The reaction mix was then made 50 mM NaCl, 50 mM Tris-hydrochloride (pH 7.5), 10 mM MgCl2, and 1 mM dithiothreitol in a final volume of 0.05 ml. Restriction enzyme (30 U) was added, and the DNA was digested at 37°C for 1 h. The free pentanucleotide generated by this digest was removed by EtOH precipitation with ammonium acetate as described by Maxam and Gilbert (29). The restricted DNA was then ligated in 0.2 ml overnight, using the buffer described above. Samples (5, 10, and 20 μl) of the ligation reaction were used for transformation of competent E. coli MC1061 cells, as described above.

Construction of subclones for copy number analysis. The 2μ...
RESULTS

Assay for ARS function. We have investigated the sequences necessary for ARS function by in vitro mutagenesis of the ARS-containing plasmids (Fig. 1). These plasmids are derivatives of YRp5, a nonself-replicating vector that contains the entire sequence of the bacterial plasmid pBR322, for efficient replication and selection in E. coli, and the yeast URA3 gene, for selection in yeasts (35, 44). YRp12 and YRp17 contain the 1.4-kilobase (kb) EcoRI-ARS1 yeast fragment, Sc4101, and are identical except that YRp17 has lost the EcoRI site indicated in Fig. 1. Yp412/URA3 contains the 855-bp EcoRI-HindIII fragment of Sc4101 (41). These plasmids define the ARS* phenotype, as we shall use it in discussing our results. That is, they transform S. cerevisiae at high frequency (300 to 2,000 transformants per μg of DNA) and can replicate extrachromosomally. The ARS* plasmids are mitotically unstable, however, losing the plasmid at a rate of 10^-1 per cell per generation. Chromosomes are lost at a rate of 10^-3 per cell per generation for comparison (9, 18). Under selective conditions, the generation time for strain SS111 carrying an ARS* plasmid is 2.5 h as compared with 1.5 h in complete medium. Based on these properties, our initial assay for ARS function after mutagenesis was for high-frequency transformation as defined above. Mutants that did not give more than one or two transformants per μg of DNA were called ARS-. Mutants that did give transformants were further characterized by comparing parent and mutant as to the size of transformant colonies, the rate of plasmid loss in nonselective medium, and the generation time in selective medium. We have chosen in most cases to report only the generation time as a qualitative indication of ARS function, since no additional, reproducible quantitative information was obtained from plasmid loss studies. The weaker the ARS, the longer the generation time, with the most defective mutants unable to grow at all in liquid medium. Others have previously used the same criterion to characterize ARS sequences for similar reasons (23, 41, 48). Our only attempt at quantitation is the copy number analysis presented in the second section of the paper.

Bal31-generated deletions to determine the left boundary of ARS1. (The left and right designations refer to the orientation of ARS1 [Fig. 2].) To localize ARS1 we have introduced deletions or insertions of sequences in or surrounding an 11-bp consensus sequence proposed to be necessary for autonomous replication in S. cerevisiae (2, 40). The consensus sequence as refined by Broach et al. (2) is $^\mathrm{TTTTACGTAC}$ and is found within the small BglII-EcoRI fragment of Sc4101 (Fig. 1), including 1 bp of the BglII recognition site. Another sequence designated ARS1' by Stinchcomb et al. (40) and differing from the consensus by only 2 bp, TCTTGTATTTA, is also found on this fragment (position 315 to 325 of Sc4101; see Fig. 2 for an explanation of the numbering system).

A series of deletions was constructed from the EcoRI site closest to the BglII site of Sc4101 (Fig. 1). Methods are described in detail above. In summary, DNAs were digested with EcoRI and treated with the exonuclease Bal31, and EcoRI linkers were ligated to the deleted fragments. The fragments were digested with EcoRI, and the DNA was ligated to form circular molecules and used to transform E. coli. Transformants were screened for deletion mutations by analysis of plasmid DNA by gel electrophoresis. Digestion with the exonuclease Bal31 results in the deletion of pBR322 sequences as well as yeast sequences.

A map depicting nine deletion mutants, designated ΔR1 to ΔR9, from the left side of ARS1, is shown (Fig. 2A). The
FIG. 2. (A) Map of the deletions from the EcoRI sites of Yp412/URA3 and YRp17. Deletions were generated by Bal31 digestion of EcoRI-cleaved Yp412/URA3 and YRp17. Shown are the deletions in relation to the EcoRI-HindIII fragment common to both YRp17 and Yp412/URA3 (Fig. 1). YRp09ΔR1 to ΔR5 are derivatives of Yp17/URA3, and YRp09ΔR6 to ΔR9 are derivatives of YRp17. Open boxes indicate clones that have an EcoRI linker at the end of the deletion. The deletions were sized by agarose gel electrophoresis to the nearest 10 bp. HFT, high-frequency transformation. Restriction sites: R, EcoRI; B, BglII; H, HindIII; P, PstI. Numbers above or below restriction sites are the distance in nucleotides from the EcoRI site, designated position 1. Position 1 is position 1453 in reference 48. (B) Map of the Bal31-generated deletions of YRp09ΔR8. Yeast plasmid YRp09ΔR8 (A) was linearized at the synthetic EcoRI site, and deletions were generated by using Bal31. Shown are the sizes of the deletions as determined by restriction enzyme digestion and electrophoresis on acrylamide gels. All the DNA clones contain synthetic EcoRI sites. The endpoint of the deletion in ΔR88 was determined by nucleotide sequencing (29). HFT, high-frequency transformation.

deletions range from 40 to 540 bp. The first five deletions, 40 to 140 bp, were constructed with plasmid Yp412/URA3. YRp17 was used to generate longer deletions, 440 to 540 bp, to maintain an intact ß-lactamase gene and thus the ability to select for ampicillin resistance in E. coli. To obtain further deletions, a third series of mutants was constructed with plasmid YRp09ΔR8 (Fig. 2), a deletion mutant with 117 bp of yeast sequence between BglII and the endpoint of the deletion. (Sequences to the right of BglII are intact.) These deletions range in size from 25 to 126 bp and are designated ΔR81 to ΔR89 (Fig. 2B). We were somewhat surprised to find that deletions to within 18 bp of BglII and within 4 bp of the consensus sequence are still capable of high-frequency transformation. In contrast, the 126-bp deletion that removes the BglII site and an additional 11 nucleotides to the right completely abolishes ARS activity (i.e., no high-frequency transformation). Thus, deletion of the consensus makes the plasmids ARS-; and at most, 18 bp to the left of the BglII site are necessary to give high-frequency transformation.

Transformation frequencies and stabilities of deletion mutants. To examine more carefully whether there is any effect at all on the ARS function of the EcoRI deletions up to but not including the consensus the relationship between the size of the deletions and the frequency of transformation was determined. DNA from clones YRp09ΔR1 to 8 and YRp09ΔR81 to 88 was used to transform strain SS111. The frequency of transformation ranged from 195 to 707 transfor-
mants per µg of DNA, similar to the frequencies determined by Tschumper and Carbon (49) for ARS1. The number of transformants, however, does not correlate under the conditions tested with the size of the deletions (data not shown), and we conclude that these results merely reflect inherent variability in the transformation system and that there is no difference in the ability of these deletion mutants to mediate high-frequency transformation.

The stability of the plasmids was next estimated by several means. First, the doubling time of cells containing the plasmids was measured. Plasmids YRp9AR1 to 8 and YRp09AR81 to 88 all give doubling times indistinguishable from that of YRp17, 2 to 2.5 h. This is similar to the generation times found by Stinchcomb et al. (40) for the entire 850-bp EcoRI-HindIII fragment of Sc4101. In contrast, the generation time for strain SS11 carrying the defective mutant ARS plasmids that will be described below is more than 12 h. Thus again, the EcoRI deletion plasmids appear fully ARS⁺.

The stability of the plasmids was further measured by the rate of loss of plasmid from transformants. YRp17, YRp09AR8, and YRp09AR88 (see Fig. 2 for plasmid designations) were grown overnight in minimal medium without uracil. The cells were then transferred to complete minimal media (SD plus required amino acids or nucleotides), and the rate of plasmid loss was estimated by determining, after every generation, the number of cells that could grow with and without uracil. Determinations were made both by replica plating and by plating onto duplicate selective and nonselective plates with similar results. As reported by others, 10% of the cells in the culture grown under selective conditions contained ARS plasmids, and after three generations in nonselective medium, this number had dropped to 3%, i.e., the same 30% loss per generation noted by others (43). No differences in the rates of loss of the three plasmids were detected. Thus, this criterion does not allow distinction among these mutants and wild type either. It is interesting to note that the sequence, ARS⁺ (40), that differs by only 2 bp from the consensus sequence, has been deleted in plasmids YRp09AR7 to 9 and YRpAR81 to 88, suggesting that this sequence, unlike the consensus, is not essential for function. Furthermore, the nature of sequences flanking the S. cerevisiae insert, although different in each construction, shows no effect on ARS function.

Insertion, deletion, and point mutations in and around the consensus sequence. To further investigate the role of the consensus sequence in ARS function, small deletions and insertions were created at the BglII site. To generate the deletion mutants, YRp12 was digested with the restriction enzyme BglII. The DNA was treated with S1 nuclease, and the deleted fragments were circularized by blunt-end ligation. Transformants were obtained and analyzed as described for the Bα13I experiments. DNAs that had lost the BglII restriction site were subsequently characterized by DNA sequence analysis as described by Maxam and Gilbert (29) (data not shown). YRp12S9 has a 20-bp deletion, and YRp12S1 has a 11-bp deletion.

Insertion mutations were created by digesting YRp12 with BglII, the cohesive ends were filled in by using the Klenow fragment of DNA polymerase I, and the plasmid was circularized by blunt-end ligation. This procedure creates a 4-bp insertion (confirmed by DNA sequencing) at the BglII site.

A comparison of the deletion and insertion mutations is shown (Fig. 3). The sequence of YRp12 is shown in Fig. 3A; the box outlines the consensus sequence, and the arrows indicate the BglII site. The 11-bp deletion of YRp12S1 is shown in Fig. 3B. The deletion not only removed the 4-bp cohesive end but also an additional 3 to 4 bp on either side of the restriction site. Note that 4 bp of the consensus sequence have been deleted. An unexpected result of this construction is that the deletion followed by religation generates a sequence TTTTATGTTAT, which can be interpreted as having a single bp point mutation in the consensus sequence (TTTTATTTT). This plasmid is incapable of high-frequency transformation; that is, it is completely ARS⁻. The results of Kearsey (23) with the ARS near the HO gene suggest that the core of that ARS contains 3 bp to the right of the consensus essential for ARS function. We have no deletions removing any one of the last 4 bp of the BglII site and therefore do not know whether they are essential at ARS1. Therefore it is not clear whether the reconstructed site is inactive because of the deletion or the point mutation, although we favor the latter since the sequence found in the 4 bp of the BglII site (597 to 601) is not conserved at any other

![Diagram](http://mcb.asm.org/)  
**FIG. 3.** Analysis of the insertion and deletion mutations generated at the BglII site of YRp12. HFT. High-frequency transformation. (A) Sequence surrounding the BglII site of YRp12. The arrows indicate the BglII restriction enzyme cleavage pattern. The box outlines the 11-bp ARS consensus sequence of YRp12 (40). (B) YRp12S1 has an 11-bp deletion created by cleaving YRp12 with BglII and treating with S1 nuclease, followed by blunt-end ligation. The deleted sequences are diagrammed by the double lines. The box shows what remains of the consensus sequence. (C) YRp12S9 has a 19-bp deletion, constructed as in (B). The sequences deleted are diagrammed by the double lines. The box shows the remaining consensus sequence. (D) YRp12323 has a 4-bp insertion created by cleaving YRp12 with BglII and filling in the ends, followed by blunt-end ligation. This insertion creates a new restriction enzyme site ClaI; the cleavage pattern is indicated by the arrows. The consensus sequence is boxed as in (A).
ARS element and since Kearsey has shown that point mutations in the consensus sequence do give the ARS− phenotype (23). A third possible explanation of the ARS− behavior is that the spacing is changed between the consensus and the elements in domain B (see below) in YRp12S1 and that this inactivates the ARS. This is rendered less likely, however, by the fact that insertions which change spacing without altering the consensus do not inactivate the ARS (see below).

The 20-bp deletion of YRp12S9 is shown (Fig. 3C). The deletion is asymmetrical, removing 6 of the 11 bp of the consensus sequence. However, the fusion generated by the construction produces the sequence TTTATGTCTTT, in which a different point mutation of the consensus arises. This plasmid is also incapable of high-frequency transformation, i.e., completely ARS−. Shown in Fig. 3D is the 4-bp insertion mutant, YRp123′23. The insertion does not disrupt the consensus sequence but generates a ClaI site as indicated by the arrows. Unlike the deletion mutants YRp12S1 and YRp12S9, the insertion mutant YRp123′23 is ARS+. Four additional insertion mutants (identified by the loss of the BglII site) were tested for ARS function and also found to be ARS−.

The consensus sequence is necessary but not sufficient for ARS function. Although the above experiments strongly suggest that the consensus sequence is necessary for ARS function, they do not determine whether it is sufficient. To determine whether the ARS consensus sequence was sufficient, we subcloned the EcoRI-BglII fragment of YRp09AR88, containing 19 bp to the left of BglII (nucleotides 582 to 601), into the EcoRI-BamHI sites of the pBR322 sequence of YIp5 (44). The resultant plasmid, YRp09sb25, is incapable of high-frequency transformation, suggesting that the consensus is not sufficient to give high-frequency transformation. (Although abortive transformants are observed when large amounts of DNA are used in the transformation, such abortive transformants are also observed with YIp5.) We note that this subclone contains the entire 14-bp core defined at the ARS near HO by Kearsey (23). Thus this core alone is not sufficient for function at ARS1. A subclone containing only the HO core sequence, which is different at three nucleotide positions from the ARS1 site, and no leftward flanking DNA has not been studied and therefore cannot be compared.

The stability of plasmids YRp09AR1 to 88, all of which contain deletions to the left of the consensus, suggested that the additional sequences required for high-frequency transformation fall to the right of the BglII site. In fact, as we will show below, additional sequences on both sides of the consensus are necessary for the full ARS+ phenotype. We have called the consensus sequence domain A.

**Analysis of the contribution of sequences to the right of the BglII site by subcloning deletion mutations.** We have shown that the consensus sequence is necessary but not sufficient for ARS function. To investigate the additional sequence required, we constructed plasmid YRPsb54, by inserting the 51-bp Psrl-EcoRI fragment of YRp09AR88 (nucleotides 582 to 627 of Sc4101) into the 4,700-bp EcoRI-Pst1 fragment of YIp5 (44). The 51-bp Psrl-EcoRI fragment spans the BglII site; 26 bp are to the right of the BglII site, and 25 bp are to the left of the BglII site (Fig. 4). This plasmid, YRp09sb54, has the same phenotype as a second plasmid designated YRp09sb117, which we constructed by subcloning the EcoRI-Pst1 fragment from YRp09AR8 (nucleotides 484 to 627 of Sc4101), which contains 117 bp to the left of the YRp17 BglII site and the same 26 bp to the right. Both plasmids are capable of high-frequency transformation, but transformants carrying these plasmids have a reduced growth rate and a doubling time in excess of 12 h, suggesting that the plasmids are less effective replicators than the parental AR88 and AR88 that contain 238 bp to the right of BglII (Fig. 5). Thus the rightward boundary of ARS1 must be to the right of the Pst1 site, as predicted by Stinchcomb et al. (41).

To further define this boundary of ARS1 a series of deletions with Bal31 were constructed from the HindIII site of Sc4101 (Fig. 1) in plasmid YRp09AR8 (Fig. 2). The HindIII site is 211 bp to the right of the Pst1 site and 237 bp to the right of the BglII site. A map of four of these deletion mutants, ranging from 77 to 201 bp, is shown in Fig. 5. Deletions of up to 120 bp are ARS+. Two deletion mutants, −191 and −201 bp, designated H200 and H201, respectively, have a phenotype similar to the subclones YRp09sb54 and YRp09sb117, a high-frequency transformation but reduced replicative ability leading to a 12-h doubling time for the transformants, and an increased rate of loss of plasmid (>30% per generation). Thus, at least 20 bp to the right of the Pst1 site and at most an additional 70 bp, are necessary for ARS+ function. We have called this domain B (Fig. 4 and 5, nucleotide positions 601 to 647 or 697). This is a much larger region than was found by Kearsey at the HO ARS (23), where only 27 bp, at most, flanking the core on the right are relevant to ARS function.

**Reconsideration of the contribution of sequences to the left**
of the consensus. Stinchcomb et al. (40) had previously shown that plasmid YRp522, which contains the complete 601-bp EcoRI-BglII fragment found in YRp17 (Fig. 1) but lacks all of domain B, gave high-frequency transformation of yeasts but was very unstable. We found that transformants carrying YRp522 had a doubling time of 7 h (compared with 2.5 h for YRp17) in the strains we used as hosts in these experiments. YRp522, however, is more stable than YRp09bs52, the subclone containing only the consensus sequence and almost no flanking yeast DNA, and is also more stable than YRp09bs117 and YRp09bs54, subclones we constructed and which contain only 117 and 19 bp to the left of BglII (Fig. 4) and, like YRp522, are missing domain B. These latter two plasmids give transformants with doubling times longer than 12 h. Thus, in the absence of domain B, sequences more than 117 bp to the left of the consensus sequence, which could be deleted with no adverse effect in the presence of domain B, apparently contribute to ARS function. These results define a third domain, the borders of which are somewhere between the consensus and the EcoRI site at position 1. We have designated this region domain C (Fig. 5). Interestingly, it is within this region, 200 bp to the left of the consensus sequence, that the center of the replication bubbles observed during in vitro replication are observed (5). Koshland and Hartwell (C. S. Newlon, in A. Rose and J. Harrison, ed., The Yeasts, vol. 3, in press) have also shown that deletions in domain C affect ARS activity.

Functions of the various domains. Having gleaned an idea of the structural organization of ARS1, we would now like to understand the functions of each region. Although domain A is clearly required for high-frequency transformation, the contributions of domains B and C are not clear from the above analysis. In an attempt to obtain a better understanding of the function and interrelationship of the three domains, particularly domains B and C, we studied the copy numbers of the wild type and the deleted plasmids. Analysis

FIG. 5. Map of the Bal31 deletions at the HindIII site of YRp09AR8. Yeast plasmid YRp09AR8 (Fig. 2) was linearized at the HindIII site, and deletions were generated with Bal31. Shown are the size of the deletions as determined by restriction enzyme digestion and electrophoresis on polyacrylamide gels. All the DNA clones contain a synthetic HindIII site. HFT, High-frequency transformations; +, growth rate and rate of plasmid loss; −, slower growth rate and increased frequency of loss over YRp17. Domain A, open box; domain B, cross-hatched box; domain C, hatched box. Domain A is defined as the 11-bp consensus and includes position 597, the first nucleotide of the BglII restriction site. Domain B extends from position 598 to somewhere between 647 ± 5 and 707 ± 5. Domain C is defined as the region between positions 400 and 597.

FIG. 6. Construction of β-galactosidase carrying plasmids for copy number determination. Plasmid pLGSD5 contains the 2.2-kb EcoRI fragment from the 2μm plasmid (B form). the 1.1-kb yeast URA3 fragment, a 2.9-kb fragment containing the E. coli lacZ gene linked to the yeast GAL10 promoter, the origin of replication, and the β-lactamase gene of pBR322 (15). The 2μm fragment of pLGSD5 was replaced, as described in the text, by the EcoRI-HindIII fragment either from YRp7 containing the complete ARS1 region or from YRp09AR8. YRp09AR8, 6H103, and 6H200 containing the deletions around the ARS1 region described in Fig. 2 and 5. All resulting plasmids, pLGRARS1, pLGR8, pLGR88, pLGH103, and pLGH200, carry the ARS1 fragment as the EcoRI-HindIII fragments of Fig. 2 and 5, and are in the same orientation. Finally a plasmid, pLG2ARS1, was constructed that carried two copies of ARS1.
of copy number mutants has been invaluable in analyzing *E.
coli* plasmid replication mutations.

We carried out blot-hybridization analysis to compare
amounts of plasmid DNA in strains carrying plasmids having
deletions in either domains C or B as described above.
Surprisingly, copy number was the same for all plasmids,
whether or not they were deleted (data not shown because
Fig. 8 below demonstrates the same point). This is different
from the findings of Stinchcomb et al. (40), who reported a
fourfold lower copy number in a lambda plasmid containing
the 638-bp *EcoRI*-*PstI* fragment of YRpl7 (i.e., domains C,
A, and 26 bp of B) relative to lambda carrying *ARS1*. The
discrepancy could be due to the difference in the constructions
or in the method used in the two studies to determine
copy number. Both methods depend on there being no
differential recovery of linear, chromosomal DNA and plasmid
DNA in the analysis.

To corroborate general trends in copy number analysis,
a different type of copy number determination was carried out.
Just as β-galactosidase fusions have been used to study the
strengths of promoters, we used a β-galactosidase as a marker
enzyme for determining plasmid copy number. *ARS1* and the
deletions R8, R88, H103, and H200 (Fig. 2 and 5) were
subcloned into the 2μm vector pLGSD5 (15; Fig. 6). This
latter plasmid contains the *E. coli lacZ* gene fused to the
yeast *GAL10* promoter such that β-galactosidase can be
induced by galactose in the medium. For our studies, the
2μm sequences were completely replaced by the *ARS1*
DNAs, as described in the legend to Fig. 6 and above. All
resulting plasmids show stabilities similar to the parental
*YRp* derivative, as estimated by the doubling time of transformants
(Fig. 7A); that is, all plasmids show growth rates similar
to the parental *ARS1*, except the domain B mutant
H200, which has increased generation times. β-Galacto-
sidase levels were determined in strains transformed with
each plasmid and grown in medium containing galactose in
addition to glucose (Fig. 8A). The average copy number was
calculated by dividing the activity found in cultures of the
plasmid-containing cells by that obtained in the culture of
D603i. Specific activities (units per cell [×10^3]) were 45, 89,
309, 295, 34, 31, 24, 30, and 24 for i, 2i, pLGSD5, SD5/2,
ARS1, R8, R88, H103, and H200, respectively. The number
obtained by comparing D603i (i) and D603i (2i) agrees well
with the expected copy number of 2 for the latter strain. The
copy number per plasmid-containing cell (Fig. 8A) was
determined by dividing the average copy number by the
percent cells containing plasmid. This was 86, 56, 3.8, 3.5,
2.0, 3.3, and 2.9% for pLGSD5, pLGSD5/2, pLGARS1,
PLGR8, PLGR88, PLGH103, and PLGH200. Note that
plasmid 2ARS1 fortuitously integrated into the chromosome
in this experiment. When not integrated, the copy number is
equal to that of *ARS1*. In summary, we find that the copy
numbers for all *ARS* plasmids, deleted or not, are similar,
20 to 30, and deleted plasmids do not have low copy numbers.
No correction has been made for the fact that cells that have
lost the plasmid may contain β-galactosidase and *URA3* gene
product activity for several generations after plasmid loss,
which introduces some error into the calculation. *GAL4*
levels do not appear to be limiting, at least up to a copy
number of 30 (also see reference 15). These assumptions do

\[ \text{FIG. 7. Growth rate of strains carrying β-galactosidase plasmids. Derivatives of plasmid pLGSD5 containing the entire } ARS1, \text{ designated } ARS1, \text{ and deletions R8, R88, H103, and H200 (see Fig. 2, 5, and 6) were introduced into strain D603, a } \text{ura3-52 strain that does not carry out glucose repression (regL). pLGSD5 and pLGSD5/2 are the 2μm-containing parental plasmids, the latter plasmid being a dimer. Growth rates of various plasmid-containing strains are expressed relative to the growth rate of strain D603 that was isolated after spontaneous integration of plasmid pLG2ARS1. D603i and D603i, designated i and 2i in the figure, contain one and two copies of integrated plasmid DNA, respectively. Numbers within parentheses refer to the percent plasmid-containing cells in the culture. Strain D603 is } [\text{cir}^+]\text{, accounting for the greater stabilities of pLGSD5 and pLGSD5/2 relative to the } ARS1 \text{ plasmids.} \]
from Fig. 7 it is clear that expression of β-galactosidase alters the growth rate of plasmid-containing cells and might have an effect on the copy number. We used the Southern blotting technique as described above to quantitate plasmid levels. The average copy number was determined as the amount of plasmid DNA divided by the sum of the amount of the two URA3 bands and the GAL10 band. Copy number per cell was calculated by dividing the average copy number by percent plasmid-containing cells. The percent plasmid-containing cells was 97% for pLGSD5, 80% for pLGSD2, 11% for pLGARS1, 9% for pLGR8, 5.5% for pLGR88, 12% for pLGH103, notably only 0.8% for pLGH200 the unstable mutant, and 11% for pLG2ARS1. Calculated copy numbers for these plasmids are 11, 21, 100, 100, 200, 200, 100, 600, and 100. Each plasmid has a higher absolute copy number when cells are grown under these conditions than when cells are grown in the presence of galactose (Fig. 7A), but relative numbers are consistent with the enzymatic method of copy number determination. This is consistent with the higher growth rate of the cells grown on glucose alone, seen by comparing Fig. 7A and B. We again conclude that the deletion mutants, whether they have a doubling time of 2.5 or 12 h, all have high copy numbers similar to that of the wild type. In fact, the least stable plasmid, H200, has an even higher copy number per cell than does the wild type.

To explain these findings we propose that the instabilities of the mutant plasmids are due to segregation defects in addition to (or rather than) replication defects. Although plasmids like 2μm segregate equal amounts of DNA into the mother and the bud (symmetric segregation), Murray and Szostak (33) recently showed that 50% of cells containing ARS plasmids segregate asymmetrically, with the entire plasmid content segregating 19:1 into the mother instead of the bud. The deletion mutations may increase the asymmetry. Clearly, a more unambiguous assay will be required to measure replication alone. One can envision such assays using centromeres to stabilize plasmids in combination with our β-galactosidase assays or using in vitro assays based on soluble replication systems as have been used in bacterial studies (see below).

**DISCUSSION**

Stinchcomb et al. (40) distinguished two functional domains within ARS1 and first proposed the importance of the consensus sequence that was later more completely defined by Broach et al. (2). We have further characterized the sequences necessary for ARS1 function in *S. cerevisiae* in vivo and have revised the description of the sequence to include three domains. The information necessary for stable autonomous replication resides in a 124-bp region. The boundaries of this region are 15 bp to the left of the BglII site, domain A, and 109 bp to the right of the BglII site, domain B. Additional sequences are necessary in the absence of domain B. These fall within domain C, which extends at least 200 bp to the left of domain A. Domain A appears to be the only absolutely essential region, since deletions of domains B or C destabilize the plasmids but do not abolish replication, as do deletions or point mutations in domain A.

A potentially useful outcome of our mutagenesis is that a plasmid was generated that contains the consensus sequence bearing a single point mutation. First, because this abolishes ARS function, it suggests that the consensus sequence is essential for high-frequency transformation of yeasts (see below for alternative interpretations of these mutations).
Furthermore, it offers the possibility of a genetic approach to isolating mutants in putative proteins that might recognize or bind to or mediate their action through the consensus sequence. One could simply select for suppressors of the point mutation that allow the plasmids to multiply again. This genetic approach would form a perfect complement to a biochemical approach to isolating these protein components of the replication apparatus.

From results obtained in this and other systems, a working model for the functional organization of ARS can be proposed. In this scheme, domain A, the consensus sequence, would be the recognition site for a DNA binding protein with a key role in replication, analogous to the DNA protein of E. coli. The DNA protein binds to a 9-bp recognition sequence within oriC and participates directly in the initiation of DNA synthesis (13). In support of our proposal, a high-molecular-weight replication complex has been shown to bind in the region of the consensus sequence by electron microscopy (22). In addition to its role in replication, the yeast consensus sequence seems to have a role in regulating histone biosynthesis and the transcriptional state of HML and HMR loci (43). Similarly, the recognition site for DNA protein is found in the promoter region of the dnaA gene, perhaps accounting for the observed autoregulation of DNA protein synthesis (see reference 13). Thus the DNA protein may be bifunctional, having a role in both transcription and replication. Although the consensus sequences found near HML, HMR, and the H2b gene could mediate their effect through replication and alteration of the state of the chromosome, they could also be directly important as recognition signals in the transcriptional processes they affect.

Domain B, which shows little if any homology between different ARSs, could contain sequences for specific regulation, such as activation of different replication origins at different times during each S phase. Since domain B apparently stabilizes the plasmids but does not affect their copy number, an additional function might be to provide a site for nuclear attachment, either to the chromosome or to the nuclear matrix or membrane. Both physical and genetic evidence suggest that the endogenous plasmid of yeast, the 2μm circle, is associated with chromosomes during portions of the cell cycle (36, 47). Murray and Szostak (33) have demonstrated that circular ARS plasmids segregate asymmetrically, suggesting that they do not diffuse freely in the nucleus. It has been proposed that the 2μm REP loci form a nuclear attachment system and that the 2μm plasmid is stable because this system somehow ensures detachment for proper segregation as necessary (24). If domain B has some attachment function, domain B deletions might be expected to function normally in the presence of a centromere on the plasmid or with the 2μm REP loci present in cis. Preliminary evidence from L. Hartwell suggests that this is in fact the case (personal communication).

A role for domain C is indicated by in vitro replication studies that show that the replication bubbles initiate in this region (5), suggesting that, although domains A and B may be necessary for replisome assembly and for regulation, DNA synthesis may actually initiate in domain C. Since domain C has a less significant role in stability, this might seem at first unreasonable. However, there is precedent for this type of organization in that replication initiates in vitro outside the minimal sequence required for oriC function in E. coli (46) and any adjacent sequence will serve as an initiation site of bidirectional replication, as long as the minimal sequence is intact. Since domain C can be replaced by E. coli plasmid DNA in the presence of domain B, with no apparent deleterious effect, the situation prevailing at ARS1 may be similar to that at oriC.

Further comparison to oriC is a useful way to analyze ARS function. oriC contains a 245-bp region required for origin function, whereas the complete ARS1 is ca. 300 bp. The transcriptional and translational signals present have been described by others (19, 39, 48). Relevant to our new results is the fact that in vitro mutagenesis and taxonomic comparisons suggest that oriC is composed of protein recognition sequences precisely spaced by intervening regions in which base substitutions do not inactivate oriC functions but even small insertions or deletions do. Single base substitutions in the protein recognition sites weaken but do not completely inactivate oriC (13, 46). As discussed above, ARS1 contains a consensus sequence that may be a protein recognition sequence. However, we have demonstrated that ARS1 differs from oriC in that small insertions between this site and the rest of the ARS do not lead to inactivation. We do not know whether deletions between the two can also be tolerated.

Therefore, just as in eucaryotic promoters (14, 30), deletions may be less or more stringent than for replication recognition in eucaryotes than in procaryotes. Furthermore, point mutations in the yeast consensus completely inactivate the HO ARS (23) and, from our work with reservations in interpretation mentioned in the text, also ARS1, whereas no point mutation completely inactivates oriC. Unlike ARS1, the borders of oriC are clearly defined, and even small deletions from either side lead to inactivation of ori function, rather than the type of modulation of function observed in yeast ARS studies. For instance, deletion of domain B does not completely inactivate plasmids but only destabilizes them, as evidenced by the behavior of YRp522, containing only domains C and A (40; this work). Furthermore, portions of domain B give partial function, as demonstrated by a comparison of the subclone containing only the consensus sequence (sb25) with the subclone containing 26 bp of domain B in addition (sb34). YRp3b25 does not give high-frequency transformation at all, whereas sb34 does, although the sb34 plasmids are very unstable. Finally, deletion of domain C shows no effect at all under certain conditions. However, we have added centromere sequences to plasmids containing the domain C deletion, and these are far less efficient replicators than the corresponding ARS1-CENIV control plasmids (F. Srienc and J. L. Campbell, manuscript in preparation). Furthermore, cotransformation with YRp17 and YRp09ΔR88 does give rise to only YRp17 clones, suggesting, but not proving, that YRp17 does replicate or is transmitted more efficiently than the deletions (data not shown).

Our most surprising finding was that stable and unstable plasmids have similar copy numbers. Changes in stability without changes in copy number have been observed in other yeast plasmids. Kikuchi (24) has shown that mutations in the REP loci of 2μm plasmids can cause reductions in stability but not in copy number.

In vivo analyses of ARS function are difficult since chromosome transmission is measured—not just replication but also segregation—is measured. Various DNA sequences have been shown to increase the fidelity of transmission of chromosomes in yeasts: centromeres, telomeres, and the 2μm REP or FLP loci or both (43). The studies presented in this paper are an attempt to study ARS function in the absence of these, to establish a background for in vitro replication studies on the plasmids. Insofar as our studies have defined domains A, B, and C they have been useful.
However, they do not allow us to decide whether the instability we observe in our mutants is due to plasmid loss, rather than to a replication defect. In addition to their usefulness in in vitro studies, these deletion mutants will be useful in vivo systems that distinguish segregation from replication. Such systems have been developed in the laboratories of R. Davis and of L. Hartwell (C. S. Newlon, in press) with plasmids that contain centromeres in addition to ARSs.

Comparison of ARS1 and the ARS at the HO gene. Recently, another ARS, that near the yeast gene HO, has been studied by methods similar to those used here (23). Similarities and differences, some a matter of definition and some of fact, are apparent. First, the consensus sequence is essential at both sites, although at HO it is TTAAAAATTTTT, differing at the fifth nucleotide from the consensus at 12 other ARSs. However, just as at ARS1 the consensus is not sufficient in the HO ARS. This was demonstrated by the fact that deletions of bases next to the consensus at HO lead to a completely ARS− phenotype. Second, as at ARS1, deletions to the left of the ARS (domain C) do not affect ARS function as long as sequences normally present at the right are intact. Perhaps effects of domain C deletions at HO would show up in the absence of domain B, or in the presence of a centromere as we have found at ARS1, but such studies have not yet been reported. Third, a much smaller region to the right of the consensus than that at ARS1 is essential at HO. Only 27 bp to the right of the consensus are necessary for the full ARS+ phenotype at HO, whereas 47 to 109 bp (domain B) are necessary at ARS1. Fourth, we originally proposed the extent of domain A as only the 11 bp of the consensus, even though we had only deleted up to within four bases to its right since these bases are not conserved at other ARSs and we did not think it likely they were essential. However, the mutations of Kearsey actually show that deletion of one of these four nucleotides, at least at HO, gives an ARS− phenotype. This leaves open the possibility that the core ARS may actually be larger than the consensus sequence alone at ARS1, as well. Since these 4 bp have not been deleted at ARS1, an uncertainty remains as to their essentiality at ARS1. It is therefore not possible to conclude whether the point mutants in the consensus that arise as a result of our S1 deletions inactivate ARS1 because of the loss of essential nucleotides, because of a change in spacing between domains A and B, or because of the point mutation in the consensus. Kearsey elegantly shows that two point mutations in the consensus sequence do inactivate the HO ARS, so it is possible that it is the point mutations that are effective at ARS1. Even if these additional nucleotides do turn out to be essential at ARS1, they are not sufficient at ARS1, since in the subclone s2b25, we have cloned a fragment that contains the sequence in question in addition to the consensus and yet the plasmid is ARS−, giving only abortive transformants. Therefore, on their own, these additional nucleotides do not make the core functional at ARS1. In interpreting the results of deletions that come to within a few nucleotides of the consensus sequence, it should be kept in mind that others have noted an ARS+ phenotype for the 601-bp EcoRI-BglII fragment (domains C + A), which we and Stinchcomb et al. (40) find is ARS+. The point is that the context of the ARS fragment in the clones is different between the ARS− and the ARS+ cases and, as pointed out by Kearsey in his studies, sometimes deletions appear ARS− that clearly are ARS+ in another context (23). Thus, the effect of DNA, to which deletion mutations are fused, clouds precise definition of the ARS boundaries. Unfortunately, none of the point mutations at HO fall in the 3 bp outside the consensus sequence. Thus, precise delineation of the structural features of the ARS awaits purification of the proteins that recognize and interact with these elements or more sensitive assays, for instance, in the presence of a centromere.

ACKNOWLEDGMENTS

This investigation was supported by Public Health Service grant GM 25508 from the National Institutes of Health and by grants from The American Cancer Society (DV-142). The March of Dimes (1-794), The National Science Foundation (CPE-8118425), and Energy Conservation and Utilization Technology. Program of the U.S. Department of Energy. J.L.C. is the recipient of RCDA CA 00544.

LITERATURE CITED


