Isolation and Characterization of Proteins that Bind to Yeast Origins of DNA Replication

Thesis by
Kevin S. Sweder

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

1989
(Submitted August 30, 1988)

To my parents

ACKNOWLEDGMENTS

I would like to thank my research advisor, Judy Campbell, for her advice and support during my graduate career. I also thank the members of my committee, Jack Richards, Scott Emr, Sunney Chan, and Norman Davidson, for their interest and suggestions on my behalf.

Many others have contributed to an enriching environment at Caltech. All the members of the Campbell lab, past and present. Two people deserve special mention: Ambrose Jong for his countless suggestions and experimental tips, and, Martin Budd for conversations which put matters into perspective and which were sometimes downright strange. Thanks to all the members of the Parker Lab especially Keith Harshman, Greg Wiederrecht, and Joanne Topol, as well as members of the Dervan machine, John Termini and Scott Youngquist. Special thanks go to Peter Rhode for his expertise and helpful discussions.

I would like to acknowledge the support and encouragement of my mother, my father, and my wife, Diana. The special friendship of Mike and Jeanie Jones has been invaluable to me.

I thank Karen Sitney for patiently wading through the quagmire that was my thesis manuscript.

Funding was provided by the March of Dimes and a Public Health Service Training Grant from the National Institutes of Health.

ABSTRACT

Yeast chromosomes contain sequences called *ARS*s which function as origins of replication *in vitro* and *in vivo*. We have carried out a systematic deletion analysis of *ARS1*, allowing us to define three functionally distinct domains, designated A, B, and C. Domain A is a sequence of 11 to 19bp, containing the core consensus element that is required for replication. The core consensus sequence, A/TTTTATPuTTTA/T, is conserved at all *ARS*s sequenced to date. A fragment containing only element A and 8 flanking nucleotides enables autonomous replication of centromeric plasmids. These plasmids replicate very inefficiently, suggesting that flanking sequences must be important for *ARS* function. Domain B also provides important sequences needed for efficient replication. Deletion of domain B drastically increases the doubling times of transformants and reduces plasmid stability. Domain B contains a potential consensus sequence conserved at some *ARS*s which overlaps a region of bent DNA. Mutational analysis suggests this bent DNA may be important for *ARS* function. Deletion of domain C has only a slight effect on replication of plasmids carrying those deletions.

We have identified a protein called *ARS* binding factor I (ABF-I) that binds to the HMR-E *ARS* and *ARS1*. We have purified this protein to homogeneity using conventional and oligonucleotide affinity chromatography. The protein has an apparent molecular weight of 135kDa and is present at about 700 molecules per diploid cell, based on the yield of purified protein and *in situ* antibody staining. DNaseI footprinting reveals that ABF-I binds sequence-specifically to an approximately 24bp sequence

that overlaps element B at *ARS1*. This same protein binds to and protects a similar size region at the HMR-E *ARS*.

We also find evidence for another *ARS* binding protein, ABF-III, based on DNaseI footprint analysis and gel retardation assays. The protein protects approximately 22bp adjacent to the ABF-I site. There appears to be no interaction between ABF-I and ABF-III despite the proximity of their binding sites.

To address the function of ABF-I in DNA replication, we have cloned the ABF-I gene using rabbit polyclonal anti-sera and murine monoclonal antibodies against ABF-I to screen a λgt11 expression library. Four EcoRI restriction fragments were isolated which encoded proteins that were recognized by both polyclonal and monoclonal antibodies. A gene disruption can now be constructed to determine the *in vivo* function of ABF-I.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Our current understanding of eukaryotic DNA replication derives largely from the extensive characterization of the process in *E. coli* and its bacteriophages (Kornberg, 1980). The roles of the various proteins and DNA sequence requirements have been elucidated through the use of highly developed *in vitro* replication systems.

The chromosomal origin of replication in *E. coli*, *oriC*, is a unique sequence contained within a 245bp stretch that is highly conserved among the Enterobacteriaceae (Oka *et al.*, 1980; Zyskind *et al.*, 1983). *OriC* contains four 9bp recognition sites for dnaA protein and a fifth site possessing seven out of nine nucleotides of the dnaA consensus sequence (Fig. 1). The left end of *oriC* contains three tandem repeats of a 13bp sequence that is highly conserved among Enterobacteriaceae (Zyskind *et al.*, 1983). Matches to this 13bp consensus sequence are present as two tandem repeats occurring only 7bp away from a dnaA binding site in the pSC101 plasmid origin, as a single close match 17bp from the dnaA binding site in the *dnaA* promoter (Hansen *et al.*, 1982; Fuller *et al.*, 1984), and as a single close match 10bp from a dnaA binding site in the 16kd gene promoter (Buhk and Messer, 1983).

The dnaA protein was originally isolated by using transducing phage carrying the gene (Chakraborty *et al.*, 1982) or by cloning the *dnaA* gene under the control of the λpL promoter (Fuller and Kornberg, 1983). The purified dnaA protein has an Mr of 52,500Da, is active as a monomer but can aggregate into complexes with lower specific activity (Fuller and Kornberg, 1983). DnaA binds sequence-specifically to the four 9bp recognition sites. This is followed by 20 to 40 dnaA monomers binding to

each other and *oriC* in a highly cooperative manner to form a large protein-DNA complex (Fuller *et al.*, 1984).

DnaA has been shown to possess a very high binding affinity for ATP and ADP that appears to regulate its role in initiation of replication. Based on nitrocellulose binding studies with dnaA and [γ-^{32}P]ATP, a K_D value of 0.03 ± 0.01μM and an ATP:dnaA ratio of 0.48 were determined. There has been no adequate explanation for this lack of stoichiometry. Once bound, ATP is slowly hydrolyzed to ADP, which remains tightly bound, by a mechanism that is stimulated by supercoiled DNA. While both the ATP and ADP forms of dnaA protein can bind to the four 9bp recognition sites, only the ATP bound dnaA protein is active in replication. Non-hydrolyzable forms of ATP such as ATPγS can replace ATP and are active in formation of an open DNA complex and in the prepriming reaction. Therefore, it would appear ATP is an allosteric effector of dnaA.

The exact order of events leading to initiation of *E. coli* DNA replication is unclear. The most likely scheme is as follows: RNA polymerase synthesizes transcripts from various promoters around the chromosome. dnaA protein binds to four 9bp recognition sequences followed by additional dnaA proteins binding cooperatively to form a complex. Those transcripts that are within 300bp of the *oriC*-dnaA complex enable the DNA in the 13mer repeats to be opened, resulting in an activated origin. Those transcripts that are not associated with an *oriC*-dnaA complex are cleaved by RNaseH, thus eliminating non-specific initiation of replication. A helicase, dnaB, is then loaded onto the template as a dnaB:dnaC complex by protein-protein interactions with dnaA.

Loading of the dnaB:dnaC complex to form a prepriming complex requires all three 13mers to unwind despite the fact that a single 13mer is sufficient for opening duplex DNA as determined by P1 nuclease sensitivity (Bramhill and Kornberg, 1988). Within the single-stranded region of the open complex, primase can synthesize RNA primers which are then elongated by the DNA pol III holoenzyme. In the presence of gyrase, to relieve topological strain, replication can proceed around the template.

Thus, several classes of proteins are required for initiation of replication: RNA polymerase, DNA polymerase, DNA primase, initiator proteins, prepriming proteins, and helicase. The mechanism by which RNA polymerase stimulates replication initiation is the least well understood. RNA polymerase is used either as a transcriptional activator or for synthesizing primers. Since the proposed primase function has only been observed in bacteriophage T7 and ColEI RNA polymerase, which can synthesize primers on the leading strand *in vitro*, is most likely required for transcriptional activation. In prokaryotes, transcription induces an open DNA conformation, allowing access to proteins and the formation of a replication complex.

Recently the activation of DNA templates by RNA polymerase has been extensively studied in *E. coli* and bacteriophage λ. RNA polymerase is required for *in vitro* initiation only at low temperature or when high concentrations of topoisomerase I or HU protein are present. Any of these three conditions make opening of duplex DNA more difficult. Normally, the *E. coli* dnaA protein acts as an initiator protein, binding at the four repeated 9mers at the origin (*oriC*), then opening the DNA helix in an

adjacent set of 13mers (Bramhill and Kornberg, 1988; Baker and Kornberg, in press). In the case of HU inhibition of replication, dnaA is unable to open the duplex DNA within the 13mer repeats as the negative superhelical density is titrated out by the HU protein, preventing duplex opening. Transcription by RNA polymerase overcomes the inhibitory effects of HU by forming an activated R-loop structure near *oriC* which allows helicase access to the template.

Eukaryotic DNA replication has been best studied in the DNA viruses SV40 and adenovirus. Several *in vitro* replication systems have been developed which replicate the substrate DNA in an SV40 origin-dependent fashion (Ariga and Sugano, 1983; Li and Kelly, 1984; Stillman and Gluzman, 1985). It was found that only 65bp of SV40 DNA could support efficient replication both *in vitro* (Stillman *et al.*, 1985) and *in vivo* (reviewed in DePamphilis and Wassarman, 1982). This minimal origin contains the T-antigen binding site II, a portion of binding site I (which contains an inverted repeat), and an AT-rich region. The intact SV40 origin contains a complete T-antigen binding site I on the early mRNA side of *ori* and a weak T-antigen binding site III which includes the G box of the early promoter. Presence of binding site I increases DNA synthesis twofold while presence of at least three out of six G boxes increases DNA synthesis another two to three fold *in vivo*.

The situation for activation of replication origins in SV40 appears to be quite different than that for *E. coli*. The SV40 chromosome replicates in the nucleus as a 5.2 kilobase pair circular template with nucleosome structure similar to or the same as the host's. Initiation of replication in

SV40 requires a genetically defined, cis-acting viral sequence called *ori*, an 81.5 kDa sequence-specific viral protein, the large T antigen, and several host proteins. The large T-antigen binds with greatest affinity to binding site I and then to binding site II. Binding site III is usually not bound in the intact SV40 *ori*. The SV40 origin of replication contains several binding sites for transcription factors. Immediately flanking the large T antigen binding sites I and II are the SV40 early promoter and enhancers. Replication can proceed when the 72bp transcriptional enhancer repeats are deleted as long as three or more of the six G boxes within the early promoter and the large T-antigen binding sites remain. Large T-antigen has been shown to possess an ATP-dependent helicase activity that can locally unwind the DNA and enable DNA primase-DNA polymerase α to initiate synthesis. This would eliminate the need for RNA polymerase II to assist in opening the duplex DNA and could explain the α-amanitin resistance of initiation of SV40 replication. The host DNA primase-DNA polymerase α complex has been strongly implicated in the initiation process (Murakami *et al.*, 1986, 1986; Hay *et al.*, 1982, 1984; Decker *et al.*, 1986; Smale and Tjian, 1986).

If RNA polymerase is not required for activating the SV40 origin of replication, what is the nature of the dependence of SV40 replication upon transcriptional enhancers and promoters? Two models have been proposed to explain the role of transcriptional elements in SV40 *ori* function. One model is that transcription factors bind to their respective recognition sequences and interact directly with proteins binding to *ori*. DNA binding by the enhancer binding protein AP-2 is inhibited in the presence of large

T-antigen by formation of a protein:protein complex between these two proteins (Mitchell *et al.*, 1987). Alternatively, proteins binding to transcriptional elements may stabilize initiation proteins binding to *ori* by facilitating strand separation.

The second model proposes that the proteins binding to the transcriptional elements alter the local chromatin structure such that the duplex DNA is then accessible to T-antigen. TFIID has already been shown to prevent nucleosome-induced repression of the adenovirus major late promoter (Workman and Roeder, 1987). Additionally, the G box of the SV40 early promoter is responsible for nuclease hypersensitive sites in the SV40 *ori* core (Gerard *et al.*, 1985).

While the bacterial chromosome is replicated from a single origin, initiation of DNA replication in eukaryotes occurs at multiple origins of replication. Initiation at these origins is not simultaneous; individual replicons initiate at various points throughout S phase (McCarroll and Fangman, 1988) and initiation from a single origin occurs only once per S phase (Fangman *et al.*, 1983). These origins must be temporally coordinated. Thus there appears to be a level of control that is not required in prokaryotic replication. Despite this difference, a great deal of mechanistic conservation exists between eukaryotic and prokaryotic replication at the level of initiation.

Yeast is an extremely valuable organism for the investigation of eukaryotic replication. Its chromosomal organization and replication are similar to those of higher eukaryotes, it is easily transformed by recombinant DNAs, and it has a well defined genetic system. The yeast

genome is 1.35 x 10^4 kb in size and distributed among seventeen linear chromosomes, mitochondrial DNA, and an endogenous extrachromosomal element called 2μm circle. The chromosomes are packaged into nucleosomes around histones as in higher eukaryotes. The chromosomes also possess centromeres, telomeres, and multiple origins of replication.

These origins of replication are the autonomously replicating sequences (*ARS*s) that allow extrachromosomal maintenance of any colinear DNA possessing a selectable marker. *ARS*s have been shown to be used as origins of replication on plasmids *in vitro* (Celniker and Campbell, 1982) and *in vivo* (Brewer and Fangman, 1987; Huberman *et al.*, 1987). Their functional domains have been defined by deletion analysis (Celniker *et al.*, 1984; Kearsey, 1984) and transposon insertion mutagenesis (Palzkill *et al.*, 1986). The resulting model is that *ARS*s consist of three domains, A, B, and C, which allow for full replication. Domain A is an 11 to 19bp sequence containing an 11bp sequence, element A, that is conserved at all *ARSs*. It is necessary and sufficient for *ARS* function on centromeric plasmids. Domain B extends 50 to 100bp 3′ to domain A, on the T-rich strand. Domain B contains a sequence about 80bp from element A, called element B, which is recognized by a 135kDa protein (Diffley and Stillman, 1988; Sweder *et al.*, in press, 1988). Domain B is crucial for efficient replication but not sufficient for *ARS* activity. Domain C is the region where replication bubbles form *in vitro*. Domain C can be deleted with minimal effect on replication. The entire *ARS* region is very high in AT content, particularly elements A and B.

The role of RNA polymerase II in initiation of replication in the single-cell eukaryote, *Saccharomyces cerevisiae*, is not well understood. Yeast RNA polymerase II is sensitive to α-amanitin *in vivo* and *in vitro*, yet studies addressing the requirement for RNA polymerase II after the obligatory protein synthesis steps have not been done.

Yeast *ARS*s share DNA sequence motifs with replication origins of several other organisms in that an AT-rich region lies immediately adjacent to a DNA-binding protein recognition site (Sweder *et al.*, in press, 1988; Diffley and Stillman, 1988). In the case of *ARS1*, deletions or insertions can be tolerated in this AT-rich region (Celniker *et al.*, 1984; Strich *et al.*, 1986).

Similar yeast to *ARS1*, *oriC* of *E. coli* contains four 9bp dnaA binding sites as well as three AT-rich 13bp tandem repeats at its "left" end. The λ origin also possesses equivalent tandem repeats in its AT-rich region. These AT-rich sequences have been shown to be the start points for unwinding of duplex DNA in a manner that is dependent upon DNA binding of a sequence-specific initiator protein. Both the dnaA protein of *E. coli* and the O protein of bacteriophage λ possess an unwinding activity that is energy dependent (Bramhill and Kornberg, 1988; Schnos *et al.*, 1988).

The SV40 origin also contains an AT-rich region immediately flanking the large T-antigen binding site II. Point mutations within the AT-rich region result in an extreme reduction in unwinding activity (Dean *et al.*, 1987). The degree to which each point mutation affects unwinding correlates precisely with the resulting deficiency in replication both *in*

vitro and *in vivo*, indicating a close relationship between unwinding and replication.

This same relationship between unwinding and replication has been observed at the 2μm *ARS* and at the histone *H4 ARS* in yeast (Umek and Kowalski, 1987; 1988). Using mung bean nuclease to detect "open" single-stranded DNA within the supercoiled plasmids, a nuclease-sensitive site was localized to the AT-rich sequence adjacent to the core consensus sequence in both *ARS*s. The amount of nuclease sensitivity correlated well with the ability of the plasmid to transform a yeast strain to prototrophy.

Currently there are three models to describe the mechanisms by which yeast *ARS*s act as origins of replication. The first is based on the unwinding studies mentioned above. Here, an initiator protein binds to the core consensus sequence and facilitates the localized melting of duplex DNA in the flanking AT-rich region. This "open" conformation then enables other proteins of the replication machinery such as helicase and DNA primase-DNA polymerase I to enter the DNA and begin synthesis. The primary sequence of the AT-rich region is not as important as the overall ability of the sequence to unwind. Plasmids in which the AT-rich region of *H4 ARS* has been replaced by different sequences from pBR322 that readily unwind were able to transform yeast efficiently.

The second model is based upon the comparative analysis of primary sequences of several *ARS*s. Palzkill and Newlon (1988) performed deletion analysis of the chromosome III *ARS*, *C2G1*, and determined the mitotic stability of the plasmid by measuring rates of plasmid loss under both nonselective and selective conditions. Those deletions that resulted in

measurable losses in transformation efficiency occurred whenever near matches to the consensus sequence were removed. The authors conclude that *ARS*s are made up of at least one consensus sequence and multiple copies of near matches to the consensus sequence. To test their hypothesis they constructed an artificial *ARS* from synthetic oligonucleotides that contained the core consensus sequence. They found that two copies of the core consensus are sufficient for *ARS* activity and that an artificial *ARS* as efficient as a natural chromosomal *ARS* could be constructed from multiple core consensus sequences in an orientation-dependent manner.

The third model is based on the isolation of proteins that bind sequence-specifically within *ARS*s. T he deletion of sequences required for protein binding would decrease replication efficiency, thus accounting for the loss of transformation frequency and plasmid stability as one deletes in from the ends of the *ARS* sequence. We and others have purified proteins that bind within *ARS1* and the *HMR-E ARS* (Sweder *et al.*, in press, 1988; Diffley and Stillman, 1988). The sequences recognized by *ARS*-binding factor I (ABF-I), *ARS*-binding factor III (ABF-III), and General Regulatory Factor I (GRFI) at various *ARS*s examined do not share extensive homology among themselves. It is likely that there are more sequence-specific proteins of lower abundance that have not been isolated yet. A protein that binds the *ARS* core consensus sequence has yet to be isolated.

In an effort to elucidate the mechanism whereby an *ARS* is involved in initiation of replication, our laboratory set out to determine the DNA sequence requirements that enabled *ARS* plasmids to replicate. The results of this work are described in Chapter 1. Despite the fact that *ARS*s

were determined to be of a tripartite nature, there remained much sequence redundancy that could not be resolved simply by sequence analysis. Therefore we set out to isolate any proteins that might interact with *ARS* sequences. The isolation and characterization of a sequence-specific DNA-binding factor, *ARS* Binding Factor I (ABF-I), is described in Chapter 2. Another protein, *ARS* Binding Factor III (ABF-III), that interacts sequence-specifically with a site immediately adjacent to the ABF-I site is also described. In Chapter 3 I describe the generation of both polyclonal and monoclonal antibodies against ABF-I and the subsequent cloning of the gene encoding *ABF-I*.

REFERENCES

Ariga, H., and Sugano, S. (1983) *J. Virol.* **48**, 481-491

Baker, T. A., and Kornberg, A. (1988) *Cell*, in press

Bramhill, D., and Kornberg, A. (1988) *Cell* **52**, 743-755

Brewer, B. J., and Fangman, W. L. (1987) *Cell* **51**, 463-471

Buhk, H. J., and Messer, W. (1983) *Gene* **24**, 265-279

Celniker, S. E., and Campbell, J. L. (1982) *Cell* **31**, 563-573

Celniker, S. E., Sweder, K. S., Srienc, F., Bailey, J. E., and Campbell, J. L. (1984) *Mol. Cell. Biol.* **4**, 2455-2466

Chakraborty, T., Yoshinaga, K., Lother, H., and Messer, W. (1982) *EMBO J.* **1**, 1545-1549

Decker, R. S., Yamaguchi, M., Possenti, R., and DePamphilis, M. L. (1986) *Mol. Cell. Biol.* **6**, 3815-3825

DePamphilis, M. L., and Wassarman, P. M. (1982) in *Organization and Replication of Viral DNA* (Kaplan, A. S., ed) pp. 37-114, CRC Press, Boca Raton, FL

Diffley, J. F. X., and Stillman, B. W. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2120-2124

Fangman, W. L., Hice, R. H., and Chlebowicz-Sledziewska, E. (1983) *Cell* **32**, 831-838

Fuller, R. S., and Kornberg, A. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 5817-5821

Fuller, R. S., Funnell, B. E., and Kornberg, A. (1984) *Cell* **38**, 889-900

Gerard, R. D., Montelone, B. A., Walter, C. F., Innis, J. W., and Scott, W. A. (1985) *Mol. Cell. Biol.* **5**, 52-58

Hansen, F. G., Hansen, E. B., and Atlung, T. (1982) *EMBO J.* 1, 1043-1048

Hay, R. T., and DePamphilis, M. L. (1982) *Cell* **28**, 767-779

Hay, R. T., Hendrickson, E. A., and DePamphilis, M. L. (1984) *J. Mol. Biol.* **175**, 131-157

Huberman, J. A., Spotila, L. D., Nawotka, K. A., El-Assouli, S. M., and Davis, L. R. (1987) *Cell* **51**, 473-481

Jongstra, J., Reudelhuber, T. L., Oudet, P., Benoist, C., Chae, C. -B., Jeltsch, J. -M., Mathis, D. J., Chambon, P. (1984) *Nature* **307**, 708-714

Kearsey, S. (1984) *Cell* **37**, 299-307

Kornberg, A. (1980) *DNA Replication*, W. H. Freeman and Company, San Francisco, CA

Li, J. J., and Kelly, T. J. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 6973-6977

McCarroll, R. M., and Fangman, W. L. (1988) *Cell*, in press

Mitchell, P. J., Wang, C., and Tjian, R. (1987) *Cell* **50**, 847-861

Murakami, Y., Wobbe, C. R., Weissbach, L., Dean, F. B., and Hurwitz, J. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 2869-2873

Murakami, Y., Toshihiko, E., Yamada, M., Prives, C., and Hurwitz, J. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 6347-6351

Oka, A., Sugimoto, K., Takanami, M., and Hirota, Y. (1980) *Mol. Gen. Genet.* **178**, 9-20

Palzkill, T. G., Oliver, S. G., Newlon, C. S. (1986) *Nucl. Acids Res.* **14**, 6247-6264

Palzkill, T. G., and Newlon, C. S. (1988) *Cell* **53**, 441-450

Schnos, M., Zahn, K., Inman, R. B., and Blattner, F. R. (1988) *Cell* **52**, 385-395

Sekimizu, K., Bramhill, D., and Kornberg, A. (1987) *Cell* **50**, 259-265

Smale, S. T., and Tjian, R. (1986) *Mol. Cell. Biol.* **6**, 4077-4087

Stillman, B. W., and Gluzman, Y. (1985) *Mol. Cell. Biol.* **5**, 2051-2060

Stillman, B. W., Gerard, R. D., Guggenheimer, R. A., and Gluzman, Y. (1985) *EMBO J.* 4, 2933-2939

Strich, R., Wootner, M., and Scott, J. F. (1986) *Yeast* **2**, 169-178

Sweder, K. S., Rhode, P. R., and Campbell, J. L. (1988) *J. Biol. Chem.*, in press

Umek, R. M., and Kowalski, D. (1987) *Nucl. Acids Res.* **15**, 4467-4480

Umek, R. M., and Kowalski, D. (1988) *Cell* **52**, 559-567

Workman, J., and Roeder, R. G. (1987) *Cell* **51**, 613-622

Zyskind, J. W., Cleary, J. M., Brusilow, W. S. A., Harding, N. E., and Smith, D. W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1164-1168

FIGURE LEGENDS

Fig 1 **Initiation of replication from *oriC*.** (A) Structure of the *E. coli oriC*. The diagram shows the four 9bp dnaA recognition sites as boxes. There is a fifth dnaA recognition site with 7 out of 9bp of the consensus between the left-most dnaA site and the others. The large solid arrows indicate the position and orientation of the 13mer repeats. Lightly shaded arrows represent RNA polymerase promoters and small arrows are for the pSC101 rep protein binding sites. (B) Model of dnaA protein recognition and activation. dnaA bound to ATP binds to the four 9bp dnaA sites and additional dnaA aggregates to form a large complex. Under low HU concentration and higher temperature, dnaA then opens the duplex DNA within the 13mer repeats. The open complex enables a helicase, dnaB, to load onto the template as a dnaB:dnaC complex to create the prepriming complex. Figures are from Bramhill and Kornberg (1988).

Fig 2 **Physical map of *ARS1*.** The diagram shows the arrangement of domains A, B, and C to each other. Boxes indicate near matches to the 11bp core consensus sequence. Arrows indicate orientation with respect to the core consensus sequence. The ABF-I binding site is represented as ABP. This site is coincident with the 3′ end of the *TRP1* gene and a bent DNA locus. The region between the core consensus sequence and ABP is very high in AT.

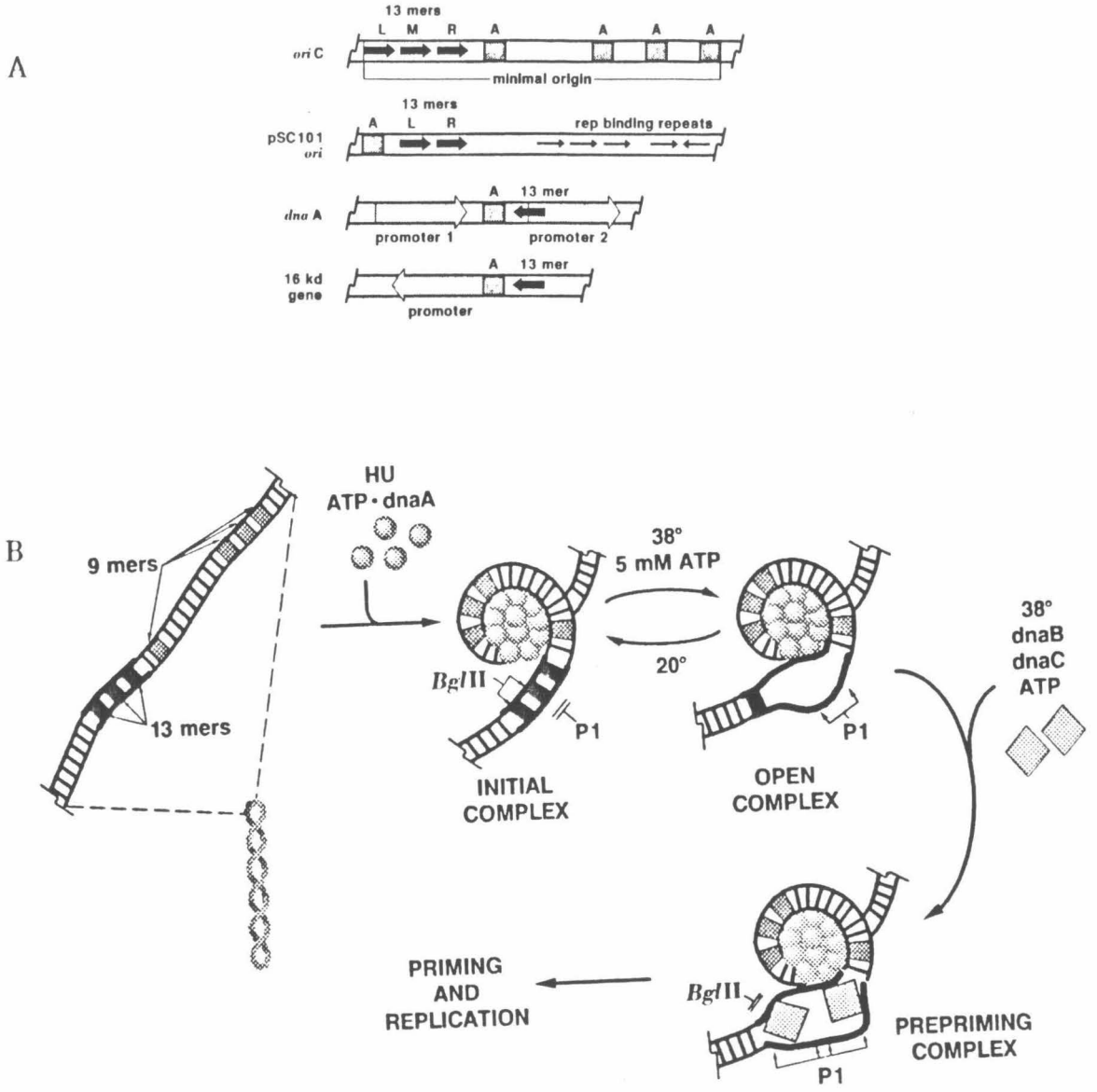

A
13 mers
L M R A A A A
oriC
minimal origin
13 mers
A L R
rep binding repeats
pSC101 ori
A 13 mer
dnaA
promoter 1
promoter 2
A 13 mer
16 kd gene
promoter
B
9 mers
13 mers
HU
ATP·dnaA
38°
5 mM ATP
20°
BglII
P1
INITIAL COMPLEX
OPEN COMPLEX
38°
dnaB
dnaC
ATP
PRIMING AND REPLICATION
PREPRIMING COMPLEX

Fig. 1.

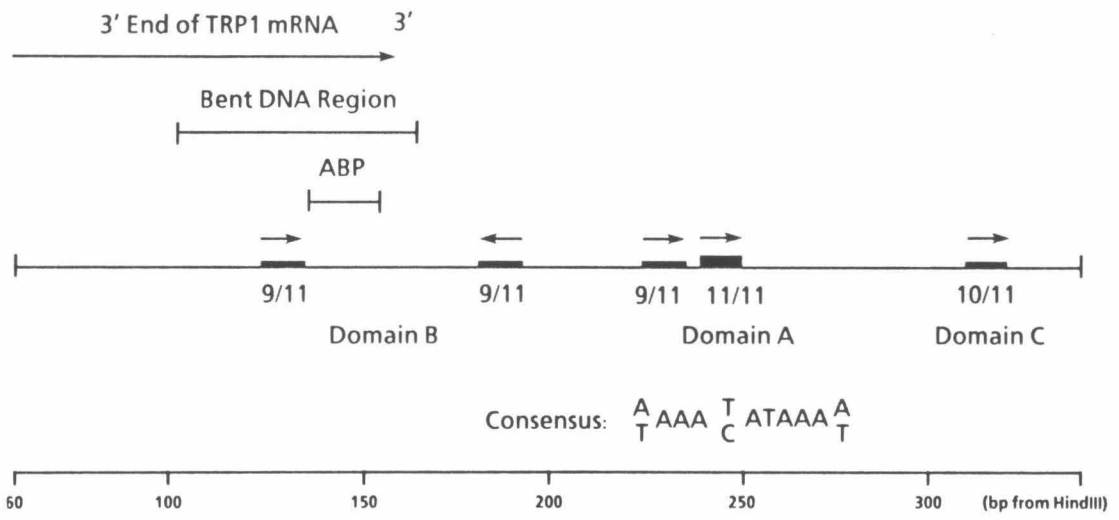

3' End of TRP1 mRNA
3'
Bent DNA Region
ABP
9/11
9/11
9/11
11/11
10/11
Domain B
Domain A
Domain C
Consensus: A/T AAA T/C ATAAA A/T
60
100
150
200
250
300
(bp from HindIII)

Fig. 2.

CHAPTER 1

Deletion Mutations Affecting Autonomously Replicating Sequence *ARS1* of *Saccharomyces cerevisiae*

Reprint from Mol. Cell. Biol. 4, 2455-2466 (1984)

MOLECULAR AND CELLULAR BIOLOGY, Nov. 1984, p. 2455–2466
0270-7306/84/112455-12$02.00/0

Vol. 4, No. 11

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 *ARS*s 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 *ARS*s, 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) *Eco*RI-*Hin*dIII 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 bisacrylylcystamine were obtained from Bio-Rad Laboratories; Sea-

* Corresponding author.
† Present address: Division of Biology, California Institute of Technology, Pasadena, CA 91125.

Kem agarose and low-melting-temperature agarose (Sea Plaque) were obtained from FMC Corp., Marine Colloids Div.; [α-^{32}P]dNTPs (400 to 3,000 Ci/mmol) were obtained from Amersham Corp.; [γ-^{32}P]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 polynucleotide kinase and T4 DNA ligase were gifts from C. C. Richardson, Harvard University, Boston, Mass. *Eco*RI and *Hin*dIII linkers were from Collaborative Research, Inc.

Strains, plasmids, and media. The *S. cerevisiae* strain used for yeast transformation, SS111 (**a** *trp1-289 ura3-1 ura3-2 his3-532 ade2-10 gal2 ino can*^{R}) was a gift from S. Scherer, California Institute of Technology, Pasadena. The yeast strain used to study the pLG plasmids (see Fig. 5) was YM603 (**a** *ura3-52 lys2-801 met his3 ade2-101 reg1-501*), which was a gift from M. Johnston, Washington University, St. Louis, Mo., and which contains a mutation (*reg1-501*) that renders the strain insensitive to catabolite repression by glucose. Strain D603 is a homozygous **a**/**a** diploid prepared from strain YM603. Plasmid pLGSD5 is described by Guarente 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 [^{32}P]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 frequencies 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*'-methylenebisacrylamide 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 Na_2–EDTA [pH 8.3]). Some DNA fragments for sequencing were isolated from acrylamide gels with the cleavable cross-linker bisacrylylcystamine 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 $ZnCl_2$, and 30 to 67 U of S1 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 $MgCl_2$, 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 S1 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 $MgCl_2$, 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 *Bal*31-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-ml reaction mix containing 60 mM NaCl, 7 mM Tris-hydrochloride (pH 7.4), 7 mM $MgCl_2$, 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 $CaCl_2$, 12 mM $MgCl_2$, 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 Bal*31 was added (27). At 1-min intervals, 10-μl samples were removed and the reaction was stopped by the addition of EGTA [ethylene glycol-bis(β-aminoethyl 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 −70°C in EtOH.

The *Bal*31-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 (*Eco*RI or *Hin*dIII) were phosphorylated in a reaction mixture containing 50 mM Tris-hydrochloride (pH 9.0), 10 mM $MgCl_2$, 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 *Bal*31-digested, dephosphorylated DNA in a 0.25-ml reaction mix which contained 50 mM Tris-hydrochloride (pH 7.5), 10 mM $MgCl_2$, 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 $MgCl_2$, 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

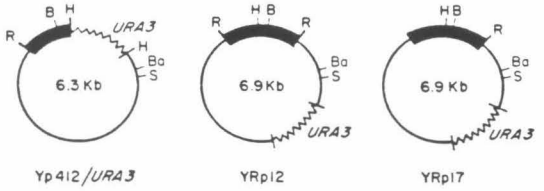


FIG. 1. Plasmids used for in vitro mutagenesis. Yp412/*URA3* is a derivative of plasmid YRp7 which contains the first identified *ARS*, *ARS1*, in the yeast fragment Sc4101 (41). Plasmid Yp412/*URA3* was constructed by digesting YRp7 with *Hin*dIII, followed by religation. This resulted in the removal of a 630-bp *Hin*dIII fragment. The *URA3* gene, contained in a 1.1-kb yeast DNA fragment, was then inserted into the *Hin*dIII site. YRp12 and YRp17 (gifts from S. Scherer) are derivatives of YIp5 and have the *ARS1* fragment Sc4101 inserted into the *Eco*RI site of pBR322 in the opposite orientation to YRp7 (35, 44). The 1.1-kb *URA3* DNA fragment was inserted into the *Ava*I site of pBR322. YRp17 has a single *Eco*RI site as indicated. Symbols: ——, pBR322 DNA; ■, yeast chromosomal sequence Sc4101 or a derivative thereof; ∼∼, yeast chromosomal DNA containing *URA3*. Restriction sites: R, *Eco*RI; B, *Bgl*II; H, *Hin*dIII; Ba, *Bam*HI; S, *Sal*I.

fragment of pLGSD5 (see Fig. 6) was replaced by the *Eco*RI-*Hin*dIII fragment either from YRp7, containing the complete *ARS1* region, or from various deletion plasmids (see Fig. 2 and 5). All constructions were carried out by the same procedure, with pLGSD5/2, which is a derivative of pLGSD5 in which we eliminated the *Hin*dIII site within the 2μ region by partially digesting pLGSD5 with *Hin*dIII and filling in and ligating the blunt ends by standard procedures. To replace 2μ DNA with *ARS1* sequences between the *Eco*RI site at the end of the 2μ sequences and the unique remaining *Hin*dIII site, pLGSD5/2 was digested partially with *Eco*RI and ligated with a complete *Eco*RI digest of YRp7 or the variants of YRp17. The ligation mixture was digested completely with *Hin*dIII and religated. To avoid a high frequency of unwanted plasmids among the transformants, the mixture was digested with *Xba*I (which has a unique site in the 2μ fragment) and *Sal*I (which has a unique site in the undesired regions of YRp7 or YRp17 plasmids) before transformation. Desired transformants were identified by the presence of a characteristic *Bgl*II-*Xho*I fragment (see Fig. 6). All resulting plasmids, pLGARS1, pLGR8, pLGR88, pLGH103, and pLGH200, carry the *ARS1* fragment as the *Eco*RI-*Hin*dIII fragments (Figs. 2 and 5) and are in the same orientation. Finally a plasmid, pLG2ARS1, was constructed that carried two copies of *ARS1*.

β-Galactosidase assay. β-Galactosidase activity was measured spectrophotometrically with whole cells after permeabilization with isopropanol as previously described (38). The reaction mixture (1 ml) contained 0.1 M sodium phosphate [pH 7.0], 0.02 M KCl, 2 mM $MgSO_4$, 25 to 200 μl of permeabilized cells, and 1 mg of *O*-nitrophenyl-β-D-galactopyranoside per ml. The reaction was carried out at 37°C and stopped by the addition of 0.3 ml of Na_2CO_3 (1 M). The suspension was centrifuged in an Eppendorf microfuge for 1 min, and the absorbance of the supernatant was measured at 420 nm. Initial rates were measured, and activity was normalized to the cell count of the permeabilized cell suspension which was determined in a Coulter Counter (Coulter Electronics, Inc.). One unit represents an increase in absorbance at 420 nm of 1.0 per min.

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 YIp5, 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) *Eco*RI *ARS1* yeast fragment, Sc4101, and are identical except that YRp17 has lost the *Eco*RI site indicated in Fig. 1. Yp412/*URA3* contains the 855-bp *Eco*RI-*Hin*dIII 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 ca. 10^{-1} per cell per generation. Chromosomes are lost at a rate of 10^{-5} 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*s for similar reasons (23, 41, 48). Our only attempt at quantitation is the copy number analysis presented in the second section of the paper.

***Bal*31-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 $^{A}_{T}$TTTATPuTTT$^{A}_{T}$ and is found within the small *Bgl*II-*Eco*RI fragment of Sc4101 (Fig. 1), including 1 bp of the *Bgl*II 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 *Eco*RI site closest to the *Bgl*II site of Sc4101 (Fig. 1). Methods are described in detail above. In summary, DNAs were digested with *Eco*RI and treated with the exonuclease *Bal*31, and *Eco*RI linkers were ligated to the deleted fragments. The fragments were digested with *Eco*RI, 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 *Bal*31 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

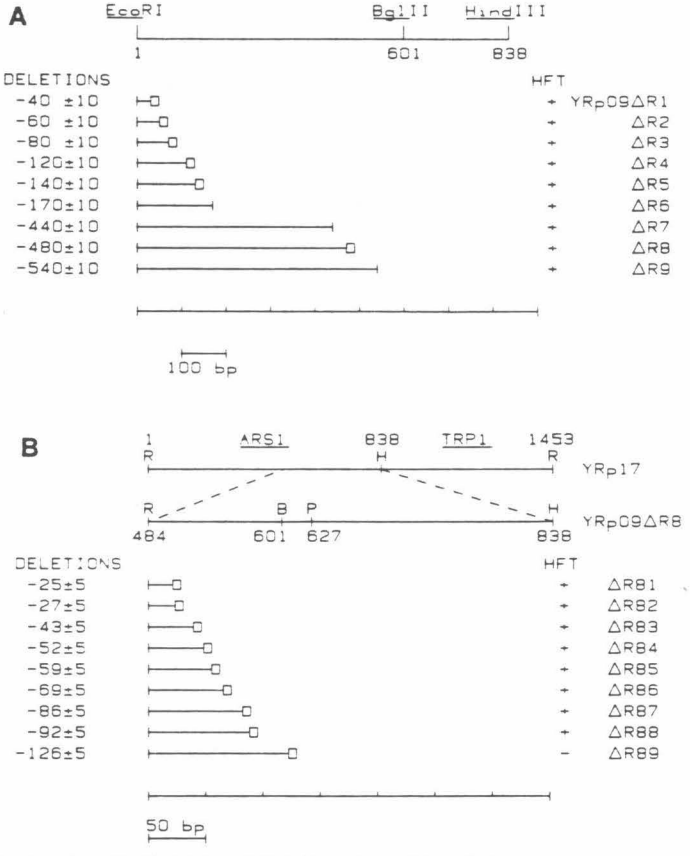


FIG. 2. (A) Map of the deletions from the *Eco*RI sites of Yp412/*URA3* and YRp17. Deletions were generated by *Bal*31 digestion of *Eco*RI-cleaved Yp412/*URA3* and YRp17. Shown are the deletions in relation to the *Eco*RI-*Hin*dIII fragment common to both YRp17 and Yp412/*URA3* (Fig. 1). YRp09ΔR1 to ΔR5 are derivatives of Yp412/*URA3*, and YRp09ΔR6 to ΔR9 are derivatives of YRp17. Open boxes indicate clones that have an *Eco*RI 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, *Eco*RI; B, *Bgl*II; H, *Hin*dIII; P, *Pst*I. Numbers above or below restriction sites are the distance in nucleotides from the *Eco*RI site, designated position 1. Position 1 is position 1453 in reference 48. (B) Map of the *Bal*31-generated deletions of YRp09ΔR8. Yeast plasmid YRp09ΔR8 (A) was linearized at the synthetic *Eco*RI site, and deletions were generated by using *Bal*31. Shown are the size of the deletions as determined by restriction enzyme digestion and electrophoresis on acrylamide gels. All the DNA clones contain synthetic *Eco*RI 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 β-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 *Bgl*II and the endpoint of the deletion. (Sequences to the right of *Bgl*II 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 *Bgl*II and within 4 bp of the consensus sequence are still capable of high-frequency transformation. In contrast, the 126-bp deletion that removes the *Bgl*II 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 *Bgl*II 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 *Eco*RI 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 YRp09ΔR1 to 8 and YRp09ΔR81 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 *Eco*RI-*Hin*dIII fragment of Sc4101. In contrast, the generation time for strain SS111 carrying the defective mutant *ARS* plasmids that will be described below is more than 12 h. Thus again, the *Eco*RI deletion plasmids appear fully ARS^+.

The stability of the plasmids was further measured by the rate of loss of plasmid from transformants. YRp17, YRp09ΔR8, and YRp09ΔR88 (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 *ARS1* 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, *ARS1'* (40), that differs by only 2 bp from the consensus sequence, has been deleted in plasmids YRp09ΔR7 to 9 and YRpΔR81 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 *Bgl*II site. To generate the deletion mutants, YRp12 was digested with the restriction enzyme *Bgl*II. 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 *Bal*31 experiments. DNAs that had lost the *Bgl*II 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 *Bgl*II, 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 *Bgl*II 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 *Bgl*II 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 ($^{A}_{T}$TTTATPuTTT$^{A}_{T}$). 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 *Bgl*II site and therefore do not know whether they are essential at *ARS1*. Therefore it is not clear whether the reconstituted 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 *Bgl*II site (597 to 601) is not conserved at any other

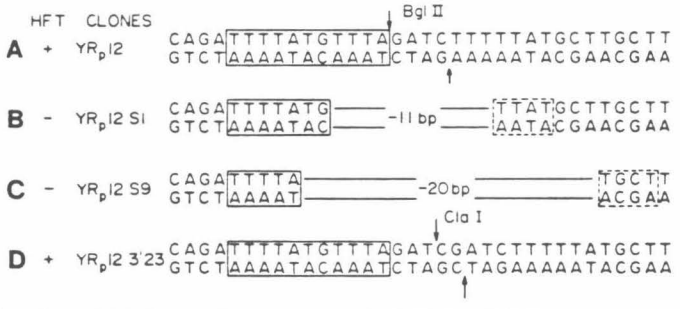
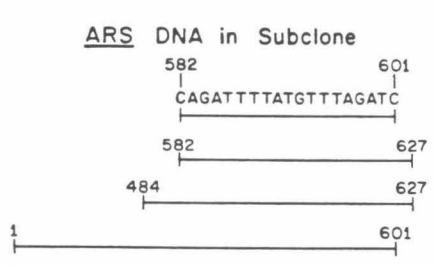


FIG. 3. Analysis of the insertion and deletion mutations generated at the *Bgl*II site of YRp12. HFT, High-frequency transformation. (A) Sequence surrounding the *Bgl*II site of YRp12. The arrows indicate the *Bgl*II restriction enzyme cleavage pattern. The box outlines the 11-bp *ARS* consensus sequence of YRp12 (40). (B) YRp12S1 has an 11-bp deletion created by cleaving YRp12 with *Bgl*II 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) YRp123'23 has a 4-bp insertion created by cleaving YRp12 with *Bgl*II and filling in the ends, followed by blunt-end ligation. This insertion creates a new restriction enzyme site *Cla*I; the cleavage pattern is indicated by the arrows. The consensus sequence is boxed as in (A).



Plasmid Designation	HFT	Generation Time (h)
YRp09sb25	−	—
YRp09sb54	+	—
YRp09sb117	+	>12
YRp522	+	7

FIG. 4. Subclones of the consensus sequence and deletion mutants. The fragments shown in YRp09sb25 and YRp522 were cloned by replacing the *Eco*RI-*Bam*HI fragment of YRp12 with the indicated fragment of YRp09ΔR88 or YRp17. The subclones sb54 and sb117 were cloned by replacing the 637-bp *Pst*I-*Eco*RI fragments of YRp12 with the indicated fragments from YRp09ΔR8 and YRp09ΔR88, respectively. The regions shown are the only yeast DNA on the plasmid (exclusively of the selectable marker, located more than 1 kb away).

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 TTTTATGCTTTT, 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 *Cla*I 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 *Bgl*II 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 *Eco*RI-*Bgl*II fragment of YRp09ΔR88, containing 19 bp to the left of *Bgl*II (nucleotides 582 to 601, Fig. 4), into the *Eco*RI-*Bam*HI 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 YRp09ΔR1 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 *Bgl*II 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 *Bgl*II 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 *Pst*I-*Eco*RI fragment of YRp09ΔR88 (nucleotides 582 to 627 of Sc4101) into the 4,700-bp *Eco*RI-*Pst*I fragment of YIp5 (44). The 51-bp *Pst*I-*Eco*RI fragment spans the *Bgl*II site; 26 bp are to the right of the *Bgl*II site, and 25 bp are to the left of the *Bgl*II site (Fig. 4). This plasmid, YRp09sb54, has the same phenotype as a second plasmid designated YRp09sb117, which we constructed by subcloning the *Eco*RI-*Pst*I fragment from YRp09ΔR8 (nucleotides 484 to 627 of Sc4101), which contains 117 bp to the left of the YRp17 *Bgl*II 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 ΔR8 and ΔR88 that contain 238 bp to the right of *Bgl*II (Fig. 5). Thus the rightward boundary of *ARS1* must be to the right of the *Pst*I site, as predicted by Stinchcomb et al. (41).

To further define this boundary of *ARS1* a series of deletions with *Bal*31 were constructed from the *Hin*dIII site of Sc4101 (Fig. 1) in plasmid YRp09ΔR8 (Fig. 2). The *Hin*dIII site is 211 bp to the right of the *Pst*I site and 237 bp to the right of the *Bgl*II 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 *Pst*I 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

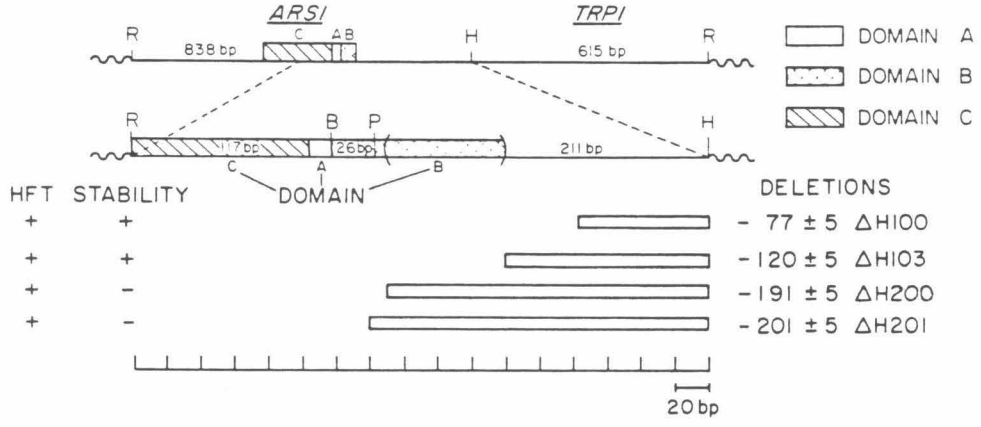


FIG. 5. Map of the *Bal*31 deletions at the *Hin*dIII site of YRp09ΔR8. Yeast plasmid YRp09ΔR8 (Fig. 2) was linearized at the *Hin*dIII site, and deletions were generated with *Bal*31. Shown are the size of the deletions as determined by restriction enzyme digestion and electrophoresis on polyacrylamide gels. All the DNA clones contain a synthetic *Hin*dIII 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 *Bgl*II 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.

of the consensus. Stinchcomb et al. (40) had previously shown that plasmid YRp522, which contains the complete 601-bp *Eco*RI-*Bgl*II 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 YRp09sb25, the subclone containing only the consensus sequence and almost no flanking yeast DNA, and is also more stable than YRp09sb117 and YRp09sb54, subclones we constructed and which contain only 117 and 19 bp to the left of *Bgl*II (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 *Eco*RI 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 *ARS1* 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

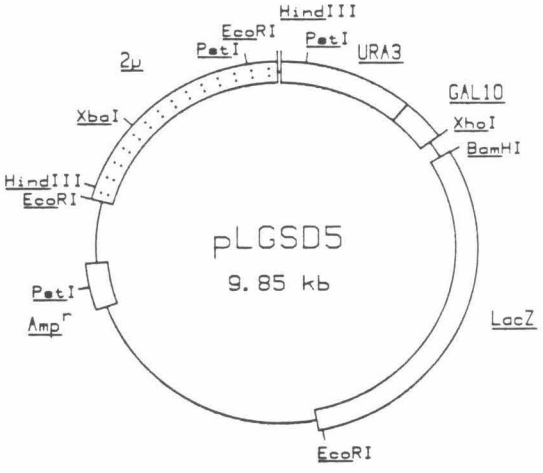


FIG. 6. Construction of β-galactosidase carrying plasmids for copy number determination. Plasmid pLGSD5 contains the 2.2-kb *Eco*RI fragment from the 2μm plasmid (B form), the 1.1-kb yeast *URA3* fragment, a 2.9-kb fragment containing the *E. coli lacZ* gene linked to the yeast *GAL10* promoter, the origin of replication, and the β-lactamase gene of pBR322 (15). The 2μm fragment of pLGSD5 were replaced, as described in the text, by the *Eco*RI-*Hin*dIII fragment either from YRp7 containing the complete *ARS1* region or from YRp09ΔR8, YRp09ΔR88, ΔH103, and ΔH200 containing the deletions around the *ARS1* region described in Fig. 2 and 5. All resulting plasmids, pLGARS1, pLGR8, pLGR88, pLGH103, and pLGH200, carry the *ARS1* fragment as the *Eco*RI-*Hin*dIII 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 mutants.

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 *Eco*RI-*Pst*I fragment of YRp17 (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 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-Galactosidase 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 [$\times 10^{-9}$]) 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 D6032i (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

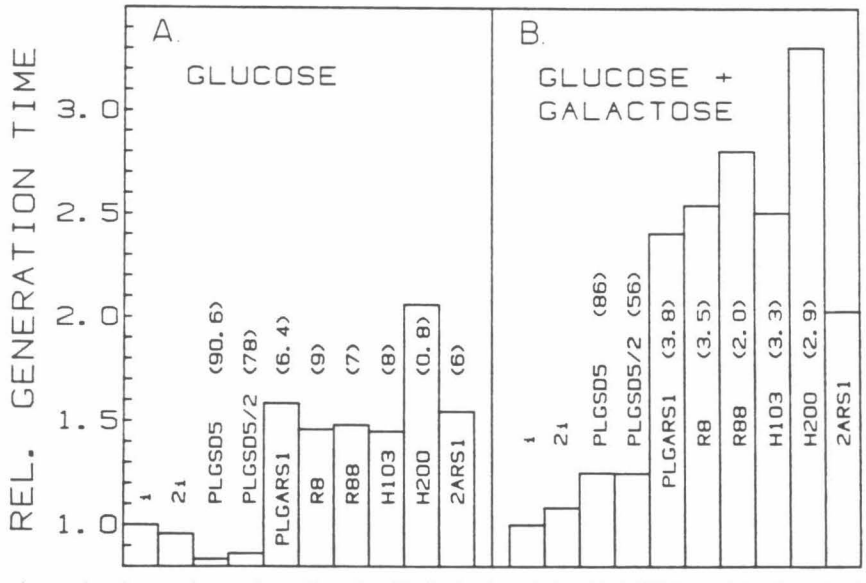


FIG. 7. Growth rate of strains carrying β-galactosidase plasmids. Derivatives of plasmid pLGSD5 containing the entire *ARS1*, designated *ARS1*, and deletions R8, R88, H103, and H200 (see Fig. 2, 5, and 6) were introduced into strain D603, a *ura3-52* strain that does not carry out glucose repression (*reg1*). 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 D603i that was isolated after spontaneous integration of plasmid pLG2ARS1. D603i and D6032i, 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 [cir$^+$], accounting for the greater stabilities of pLGSD5 and pLGSD5/2 relative to the *ARS1* plasmids.

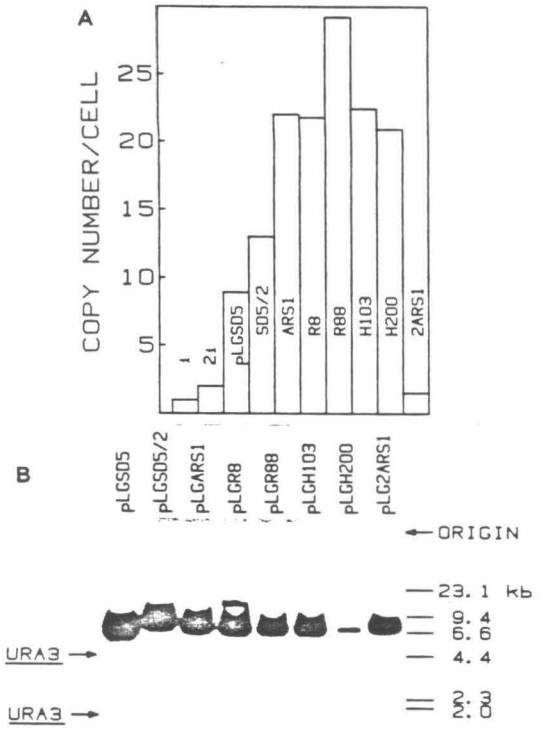


FIG. 8. (A) Copy number of deletion mutants. Cultures of strain D603 transformed with designated plasmids (50 ml) were grown in SD minimal medium lacking uracil and containing glucose and galactose (2% each) at 30°C. Doubling times were determined, and when cells reached logarithmic growth, they were plated on medium containing or lacking uracil to determine the fraction of plasmid-containing cells. A sample was taken for the determination of β-galactosidase activity, and copy number was determined as described in the text. (B) Copy number as determined by Southern blotting. Cells were grown to the logarithmic phase of growth in medium containing only glucose and harvested, and DNA was prepared as described in the text. DNA was digested with *Hin*dIII. Whereas *Hin*dIII cleaves outside of the *URA3* gene, the *ura3-52* allele in D603 contains a TY1 insertion, and digestion with *Hin*dIII gives rise to two fragments hybridizing to a *URA3* probe. *Hin*dIII cleaves most of the plasmids used in this study once. pLGSD5 and pLG2ARS1 give two fragments with *Hin*dIII, but only one is homologous to the probe. The DNA was run on an agarose gel and blotted to nitrocellulose. A nick-translated, ^{32}P-labeled DNA fragment (10^6 cpm; 1.5 kb) containing only *URA3* and *GAL10* sequences, the *Hin*dIII-*Xho*I fragment shown in Fig. 6, was added. After hybridization and autoradiography, radioactivity in individual bands was determined by densitometer analysis. (Longer exposures were used to estimate the amount of radioactivity in the *URA3* bands.) Calculation of copy number is presented in the text.

not, in any case, affect our conclusions about relative copy numbers.

Copy number was also determined in cultures grown on glucose such that β-galactosidase is not expressed, since 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 pLGSD/2, 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 *Bgl*II site, domain A, and 109 bp to the right of the *Bgl*II 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 *ARS1* 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 dnaA protein of *E. coli*. The dnaA 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 dnaA protein is found in the promoter region of the dnaA gene, perhaps accounting for the observed autoregulation of dnaA protein synthesis (see reference 13). Thus the dnaA 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 *ARS*s, 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), spacing may be less important in replicator 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 (sb54). YRpsb25 does not give high-frequency transformation at all, whereas sb54 does, although the sb54 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 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 *ARS*s.

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 TTTTAATATTTT, differing at the fifth nucleotide from the consensus at 12 other *ARS*s. 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 *ARS*s 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 sb25, 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 *Eco*RI-*Bgl*II 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 (MV-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

1. **Beach, D., M. Piper, and S. Shall.** 1980. Isolation of chromosomal origins of replication in yeast. Nature (London) **284**:185–187.
2. **Broach, J., Y. Li, J. Feldman, M. Jayaram, J. Abraham, K. Nasmyth, and J. Hicks.** 1983. Localization and sequence analysis of yeast origins of DNA replication. Cold Spring Harbor Symp. Quant. Biol. **47**:1165–1173.
3. **Campbell, J. L.** 1983. Yeast DNA replication in vitro and in vivo, p. 109–155. *In* J. Setlow and A. Hollaender (ed.), Genetic engineering, vol. 5. Plenum Publishing Corp., New York.
4. **Casadaban, M., and S. Cohen.** 1980. Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. J. Mol. Biol. **138**:179–207.
5. **Celniker, S., and J. Campbell.** 1982. Yeast DNA replication in vitro: initiation and elongation events mimic in vivo processes. Cell **31**:563–573.
6. **Chan, C., G. Maine, and B.-K. Tye.** 1981. Functional components of the *Saccharomyces cerevisiae* chromosomes. Replication origins and centromeric sequences, p. 455–472. *In* D. S. Ray (ed.), The initiation of DNA replication. Academic Press, Inc., New York.
7. **Chan, C., and B.-K. Tye.** 1980. Functional components of the *Saccharomyces cerevisiae* chromosomes. Proc. Natl. Acad. Sci. U.S.A. **77**:6329–6333.
8. **Chan, C., and B.-K. Tye.** 1983. Organization of DNA sequences and replication origins at yeast telomeres. Cell **33**:563–573.
9. **Clarke, L., and J. Carbon.** 1983. Genomic substitutions of centromeres in *Saccharomyces cerevisiae*. Nature (London) **305**:23–28.
10. **Conrad, S., and J. Campbell.** 1979. Characterization of an improved in vitro DNA replication system for *Escherichia coli* plasmids. Nucleic Acids Res. **6**:3289–3034.
11. **Fangman, W. L., R. Hice, and E. Chlebowicz-Sledziewska.** 1983. ARS replication during the yeast S phase. Cell **32**:831–838.
12. **Favalaro, J., R. Treisman, and R. Kamen.** 1980. Transcription maps of polyoma virus specific RNA: analysis by two-dimensional nuclease S1 gel mapping. Methods Enzymol. **65**:718–749.
13. **Fuller, R., and A. Kornberg.** 1983. Purified dna A protein in initiation of replication at the *E. coli* chromosomal origin of replication. Proc. Natl. Acad. Sci. U.S.A. **80**:5817–5821.
14. **Guarente, L.** 1984. Yeast promoters: positive and negative elements. Cell **36**:799–800.
15. **Guarente, L., R. Yocum, and F. Gilford.** 1982. A *GAL10-CYC1* hybrid yeast promoter identifies the *GAL 4* regulatory region as an upstream site. Proc. Natl. Acad. Sci. U.S.A. **79**:7410–7414.
16. **Hansen, J.** 1977. Electrophoresis of ribonucleic acid on a polyacrylamide gel which contains disulfide crosslinkages. Anal. Biochem. **76**:37–44.
17. **Harland, R. M., and R. Laskey.** 1980. Regulated replication of DNA microinjected into eggs of *Xenopus laevis*. Cell **21**:761–771.
18. **Hartwell, L. H., S. K. Dutcher, J. S. Wood, and B. Garvik.** 1982. The fidelity of mitotic chromosome reproduction in *S. cerevisiae*, p. 28–38. *In* Berkeley workshop on recent advances in yeast molecular biology: recombinant DNA. U.S. Department of Energy, Washington D.C.

19. **Hirota, Y., A. Oka, K. Sugimoto, K. Asada, H. Sasaki, and M. Takanami.** 1981. *Escherichia coli* origin of replication: structural organization of the region essential for autonomous replication and the recognition frame model, ICN-UCLA Symp. Mol. Cell. Biol. **22**:1–12.
20. **Hsiao, C., and J. Carbon.** 1979. High-frequency transformation of yeast by plasmids containing the cloned *ARG4* gene. Proc. Natl. Acad. Sci. U.S.A. **76**:3829–3833.
21. **Ito, H., Y. Fukuda, K. Murata, and A. Kimura.** 1983. Transformation of intact yeast cells treated with alkaline cations. J. Bacteriol. **153**:163–168.
22. **Jazwinski, S. M., A. Niedzwiecka, and G. Edelman.** 1983. In vitro association of a replication complex with a yeast chromosomal replicator. J. Biol. Chem. **258**:2754–2757.
23. **Kearsey, S.** 1984. Structural requirements for the function of a yeast chromosomal replicator. Cell **37**:299–307.
24. **Kikuchi, Y.** 1983. Yeast plasmid requires a *cis*-acting locus and two plasmid proteins for its stable maintenance. Cell **35**:487–493.
25. **Kiss, G., A. Aim, and R. Pearlman.** 1981. Regions of *Tetrahymena* rDNA allowing autonomous replication of plasmids in yeast, p. 607–614. *In* D. S. Ray (ed.), The initiation of DNA replication. Academic Press, Inc., New York.
26. **Kuo, C.-L., and J. L. Campbell.** 1983. Cloning of *Saccharomyces cerevisiae* DNA replication genes: Isolation of the *CDC8* gene and two genes that compensate for the *cdc8-1* mutation. Mol. Cell. Biol. **3**:1730–1737.
27. **Legerski, R., J. Hodnet, and H. Gray, Jr.** 1978. Extranuclear nucleases of *Pseudomonas* Bal 31. III. Use of the double-strand deoxyribonuclease activity as the basis of a convenient method for the mapping of fragments of DNA produced by cleavage with restriction enzymes. Nucleic Acids Res. **5**:1445–1464.
28. **Maniatis, T., E. Frisch, and J. Sambrook.** 1982. Molecular cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
29. **Maxam, A., and W. Gilbert.** 1980. Sequencing end-labelled DNA with base specific chemical cleavages. Methods Enzymol. **65**:499–560.
30. **McKnight, S.** 1982. Functional relationships between transcriptional control signals of the thymidine kinase gene of herpes simplex virus. Cell **31**:355–365.
31. **Montiel, J., C. Norbury, M. Tuite, M. Dodson, J. Mills, A. Kingsman, and S. Kingsman.** 1984. Characterization of human chromosomal DNA sequences with replicate autonomously in *Saccharomyces cerevisiae*. Nucleic Acids Res. **12**:1049–1068.
32. **Mullins, J., J. Casey, M. Nicolson, K. Burk, and N. Davidson.** 1981. Sequence arrangement and biological activity of cloned feline leukemia virus proviruses from a virus-productive human cell line. J. Virol. **38**:688–703.
33. **Murray, A., and J. Szostak.** 1983. Pedigree analysis of plasmid segregation in yeast. Cell **34**:961–970.
34. **Nasmyth, K., and S. Reed.** 1980. Isolation of genes by complementation in yeast: molecular cloning of a cell-cycle gene. Proc. Natl. Acad. Sci. U.S.A. **77**:2119–2123.
35. **Scherer, S., and R. Davis.** 1979. Replacement of chromosome segments with altered DNA sequences constructed in vitro. Proc. Natl. Acad. Sci. U.S.A. **76**:4951–4955.
36. **Sigurdson, D. C., M. E. Gaarder, and D. M. Livingston.** 1981. Characterization of the transmission during cytoductant formation of the 2 μm DNA plasmid from saccharomyces. Mol. Gen. Genet. **183**:59–65.
37. **Southern, E.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
38. **Srienc, F., J. Campbell, and J. Bailey.** 1983. Detection of bacterial beta-galactosidase activity in individual *Saccharomyces cerevisiae* cells by flow cytometry. Biotechnol. Lett. **5**:43–48.
39. **Stinchcomb, D., C. Mann, and R. Davis.** 1982. Centromeric DNA from *Saccharomyces cerevisiae*. J. Mol. Biol. **158**:157–179.
40. **Stinchcomb, D., C. Mann, E. Selker, and R. Davis.** 1981. DNA sequences that allow the replication and segregation of yeast chromosomes, p. 473–488. *In* D. S. Ray (ed.), The initiation of DNA replication and segregation of yeast chromosomes. Academic Press, Inc., New York.
41. **Stinchcomb, D., K. Struhl, and R. Davis.** 1979. Isolation and characterization of a yeast chromosomal replicator. Nature (London) **282**:39–43.
42. **Stinchcomb, D., M. Thomas, J. Kelly, E. Selker, and R. Davis.** 1980. Eukaryotic DNA segments capable of autonomous replication in yeast. Proc. Natl. Acad. Sci. U.S.A. **77**:4559–4563.
43. **Struhl, K.** 1983. The new yeast genetics. Nature (London) **305**:391–397.
44. **Struhl, K., D. Stinchcomb, S. Sherer, and R. Davis.** 1979. High frequency transformation of yeast: autonomous replication of hybrid DNA molecules. Proc. Natl. Acad. Sci. U.S.A. **76**:1035–1039.
45. **Szostak, J., and E. Blackburn.** 1982. Cloning yeast telomeres on linear plasmid vectors. Cell **29**:245–255.
46. **Takanami, M., S. Tabata, A. Oka, K. Sugimoto, H. Sasaki, S. Yasuda, and Y. Hirota.** 1983. The *Escherichia coli* origin of replication: essential structure for bidirectional replication, p. 257–273. *In* N. Cozzarelli (ed.), Mechanisms of DNA replication and recombination. Alan R. Liss, Inc., New York.
47. **Taketo, M., S. Jazwinski, and G. Edelman.** 1980. Association of the 2 μm DNA plasmid with yeast folded chromosomes. Proc. Natl. Acad. Sci. U.S.A. **77**:3144–3148.
48. **Tschumper, G., and J. Carbon.** 1980. Sequence of a yeast DNA fragment containing the chromosomal replicator and the *TRP1* gene. Gene **10**:157–166.
49. **Tschumper, G., and J. Carbon.** 1982. Delta sequences and double symmetry in a yeast chromosomal replicator region. J. Mol. Biol. **156**:293–307.
50. **Vogel, M., and D. Bonner.** 1956. Acetylornithase of *Escherichia coli*: partial purification and some properties. J. Biol. Chem. **218**:97–106.
51. **Zakian, V.** 1981. Origin of replication from *Xenopus laevis* mitochondrial DNA promotes high frequency transformation of yeast. Proc. Natl. Acad. Sci. U.S.A. **78**:3128–3132.
52. **Zakian, V. A., and J. F. Scott.** 1982. Construction, replication, and chromatin structure of TRP1 RI circle, a multiple-copy synthetic plasmid derived from *Saccharomyces cerevisiae* chromosomal DNA. Mol. Cell. Biol. **2**:221–232.

CHAPTER 2

Purification and Characterization of Proteins That Bind To Yeast *ARS*s

J. Biol. Chem. (1988) accepted for publication.

ABSTRACT

Two proteins that bind to yeast *ARS* DNA have been purified using conventional and oligonucleotide affinity chromatography. One has been purified to homogeneity and has a mass of 135kDa. Competitive binding studies and DNase I footprinting show that the protein binds to a sequence about 80bp away from the core consensus in the region known as domain B. This region has previously been shown to be required for efficient replication of plasmids carrying *ARS1* elements. To further investigate whether the protein might have a function related to the ability of *ARS*s to act as replicators, binding to another *ARS* was tested. The protein binds to the functional *ARS* adjacent to the silent mating type locus HMR, called the HMR-E *ARS*, about 60 base pairs from the core consensus sequence. Surprisingly, there is little homology between the binding site at the HMR-E *ARS* and the binding site at *ARS1*. The 135kDa protein is probably the same as ABF-I (SBF I) (Shore, D. *et al.*, 1987; Buchman *et al.*, 1988).

A second DNA binding protein was separated from ABF-I during later stages of the purification. This protein, which we designate ABF-III, also binds specifically to the *ARS1* sequence, as shown by DNase I footprinting, at a site adjacent to the ABF-I recognition site. Purification of these two *ARS* binding proteins should aid in our understanding of the complex mechanisms that regulate eukaryotic DNA replication.

INTRODUCTION

Yeast *ARS*s allow autonomous replication of colinear DNA when cloned into a plasmid containing a yeast selectable marker (Struhl *et al.*, 1979). Physical mapping of replication intermediates formed *in vivo* or *in vitro* suggests that *ARS*s serve as origins of replication on *ARS*-containing plasmids (Brewer and Fangman, 1987; Celniker and Campbell, 1982; Huberman *et al.*, 1987; Newlon *et al.*, 1981). That *ARS*s are also chromosomal origins of replication is implied by: 1) analogy with bacteria (Yasuda and Hirota, 1977), 2) the fact that the number of *ARS*s in yeast corresponds with the number of replicons (Chan and Tye, 1980; Petes and Williamson, 1975; Rivin and Fangman, 1980), and 3) the coincidence of replication bubbles mapped by electron microscopy and functional *ARS*s mapped in the spacer regions of the ribosomal DNA repeats (Saffer and Miller, 1986).

We and others have carried out extensive mutational analysis of *ARS1* (Celniker *et al.*, 1984), the 2μm *ARS* (Broach *et al.*, 1983), the HO *ARS* (Kearsey, 1984), the *ARS* at the HMR-E locus (Abraham *et al.*, 1984; Brand *et al.*, 1985), the H4 *ARS* (Bouton and Smith, 1986) and *ARS*s from Drosophila and human DNA (Marunouchi *et al.*, 1987; Montiel *et al.*, 1984). *ARS1* can be divided into three functionally distinct domains, designated A, B and C, as shown in Fig. 1. Domain A is a short stretch of nucleotides, 11 to 19bp, located between the HindIII site and domain C. This domain contains a conserved 11bp element, A/T TTTATPuTTT A/T, usually called the core consensus (Broach *et al.*, 1983; Stinchcomb *et al.*, 1981). ("Domain" is used here to denote a general region defined by

deletion mutations, while "element" is used to define a specific sequence of nucleotides defined by point mutations, sequence conservation or protein binding). Point mutations in the core consensus cause complete loss of *ARS* function, defining this as an essential element (Celniker *et al.*, 1984; Kearsey *et al.*, 1984). A 19bp segment of *ARS1* containing only the core consensus and 8 flanking nucleotides allows autonomous replication of centromere-containing plasmids. However, replication is very inefficient, suggesting that flanking DNA is critical (Srienc *et al.*, 1985). It is likely that domain B, an AT-rich region which extends 50 to 100bp 3′ to the core consensus as defined in Fig. 1, provides the most important elements for *ARS* function outside of domain A. Domain B- plasmids transform at high frequency but the doubling time of transformants is 7 times that of wild-type since $<1\%$ of the cells in culture contain plasmid. Domain B contains multiple copies of close matches to the core consensus, which are important for *ARS* activity, at least at one *ARS* (Palzkill and Newlon, 1988). Domain B may also contain a sequence conserved at some but not all *ARS*s (Palzkill *et al.*, 1986). Finally, domain B at *ARS1* contains a region of bent DNA that may be important (Snyder *et al.*, 1986). The remaining part of the *ARS*, Domain C, is the region where replication bubbles form during *in vitro* replication of *ARS1*. Deletions of this domain cause a small but measurable defect in replication of plasmids carrying them (Celniker *et al.*, 1984; Koshland *et al.*, 1985; Srienc *et al.*, 1985). In order to understand *ARS* function, the complex protein:DNA interactions suggested by the above studies will have to be defined.

Recently, Shore *et al.* (1987) and Buchman *et al.* (1988) have described an activity in crude fractions of yeast extracts that binds to both HMR-E and to element B of *ARS1* and have suggested that binding is due to a single protein species, which they called *ARS* binding factor I (ABF-I) or silencer binding factor B (SBF-B). Since it is now possible to purify DNA binding proteins by oligonucleotide affinity chromatography, we report here the purification of ABF-I to homogeneity. Further purification of ABF-I has also been reported by Diffley and Stillman (1988). Our results differ, however, in that we describe a purification of ABF-I that is more rapid and that yields sixty times more ABF-I per gram of cells than the protocol published by Diffley and Stillman (1988). Furthermore, we have identified a novel protein, ABF-III, that binds specifically to *ARS1* adjacent to the ABF-I site.

EXPERIMENTAL PROCEDURES

Experimental Procedures are in the form of a Miniprint Supplement.

RESULTS

Purification of ARS binding proteins-Gel retardation was used after each chromatographic procedure to monitor binding of yeast proteins to a labeled DNA fragment containing both domains A and B of *ARS1* as described in "Experimental Procedures." The DNA fragments used in each experiment are indicated in the relevant figure legends. Although competition binding studies using specific oligonucleotides and DNase I footprinting were carried out at each step of the purification, the data from

such experiments are shown only for Fraction VI, where they are most relevant.

Because of previous success in isolating DNA binding proteins in this laboratory using methods based on DNA affinity chromatography as described by Alberts and Herrick (1971), we chose native DNA cellulose chromatography as our initial purification step. In order to remove endogenous nucleic acids that interfere with such chromatography by competing with the column for DNA binding proteins, DNA was removed from extracts by phase extraction with polyethylene glycol at high ionic strength. The DNA-free extract was then passed over a DNA cellulose column at an unusually high protein to DNA cellulose ratio (Table I), and the column was washed with 10 to 15 column volumes of 0.10 M KCl to remove all low affinity binding proteins. The sensitivity of the gel retardation assay allowed monitoring of the activity to assure that no specific binding protein was washed off during loading or washing. When the proteins were eluted with a salt gradient, a single protein-DNA complex was formed by proteins eluting between 0.3 and 0.4 M KCl. Fractions giving rise to the complex were pooled, and designated Fraction III. The effectiveness of the step is quantitated in Table I, which shows that Fraction III represented more than an 250-fold purification.

Chromatography on Mono Q-Fraction III was further purified by phosphocellulose chromatography and the resulting binding activity, Fraction IV, was loaded onto a Mono Q HR 5/5 column. The high resolving capabilities of FPLC separated the binding activity into two distinct fractions. The major portion of the binding activity (82%) eluted from the

Mono Q column between 0.3 and 0.4 M KCl and gave a complex comparable to that seen on DNA cellulose and phosphocellulose (Fig. 2A, fractions 22 and 23). When this activity was pooled and analyzed by DNase I footprinting, it showed a footprint identical to that obtained with homogeneous ABF-I in the experiment shown in Fig. 6.

The remaining 18% of the binding activity eluted earlier than ABF-I (Fig. 2A, fractions 18 and 19). Although this activity gave rise to a ladder of bands with a greater mobility than the ABF-I complex, surprisingly, competitive binding experiments showed that the binding activity in fractions 18 and 19 was also specific for the ABF-I binding site. Thus, these fractions either contain a degraded form of ABF-I or a related protein.

Oligonucleotide affinity chromatography-Fractions giving rise to the ladder of bands (e.g., 18 and 19 in Fig. 2A) were pooled and designated Fraction Va. Fractions containing the major portion of ABF-I (e.g., 22 and 23 in Fig. 2A) were combined to give Fraction Vb. The proteins in these fractions were then further purified using oligonucleotide affinity chromatography. A 33bp oligonucleotide, defined as element B, according to the boundaries of the footprint in Fig. 6, was synthesized and used to prepare the affinity column as described in "Experimental Procedures."

When Fraction Va was loaded onto the column and the column washed extensively with 0.1 M KCl, no binding activity was found in the pass through or wash. Binding activity eluted from the column in two well separated peaks, one at about 0.2 M KCl and a second peak at about 0.7 M KCl (Fig. 3A). The complexes present in the lower salt fraction, Fraction

VIc, appear to involve more than one protein. Because these proteins elute at the same salt concentration from random DNA cellulose and from the oligonucleotide column, we conclude that they are not specifically recognizing the sequence in the 33bp element B oligonucleotide and suggest they are either non-specific or recognize another sequence in the fragment used for the gel retardation assay. As will be shown below, Fraction VIc does contain an *ARS*-specific protein, which we call ABF-III.

The higher salt binding fractions, on the other hand, Fraction VIa, yield a ladder of faster migrating protein:DNA complexes which can be titrated down to one protein:DNA complex at lower protein concentrations and/or higher neutral competitor DNA. As will be shown below, these fractions footprint the ABF-I site; therefore, this fraction is specific for the oligonucleotide.

Fraction Vb was loaded onto a 2 ml element B oligonucleotide affinity column. Binding activity eluted at 0.2 M and at 0.7 M KCl (Fig. 3B). The first peak of binding activity was designated Fraction VIc', since it eluted from the oligonucleotide affinity column at 0.2 M salt and gave rise to a slowly migrating protein:DNA complex with the same characteristics as Fraction VIc as defined above. The differences in mobility between Fractions VIc and VIc' are probably due to differences in stoichiometry but the fractions may also contain a slightly different mixture of proteins. The second peak of activity, eluting at 0.7 M salt, gave rise to a complex identical to Fraction Vb and accounted for the major recovery of ABF-I as shown in Table I and below. The latter fraction is defined as Fraction VIb. It is important to mention that the VIc' complex migrates distinctly slower

than Fraction VIb complex, as revealed by close examination of Fig. 3 and data not shown.

Analysis of the purified proteins by gel electrophoresis-Fractions VIa, b and c were analyzed by electrophoresis in acrylamide gels containing SDS (Fig. 4). As shown in lane 1 of Fig. 4, Fraction VIb, the major ABF-I activity is essentially homogeneous after the oligonucleotide column. To insure that the 135kDa doublet seen in the silver stained gel (Fig. 4A, lane 1) represented ABF-I, binding activity was measured after renaturation of proteins separated on an identical gel and transferred to nitrocellulose ("Southwestern" blotting), as described in "Experimental Procedures." As shown in Fig. 4B, lane 1 and insert, both bands in the 135 kDa doublet are active in binding the BglII-HindIII fragment of plasmid YRp7. Specificity of binding is verified by the absence of bands in lanes 2 and 4, which contain much more protein than lanes 1 or 3.

When Fraction VIa was analyzed by SDS gel and Southwestern blotting, no protein was detected by silver staining and no binding activity was seen (Fig. 4A and B, lanes 3). This is probably because the recovery of protein in this fraction (see Table I) is too low to allow detection by silver staining or Southwestern blotting, given the amount of Fraction VIa loaded on the gel here.

Fraction VIc contains several proteins, ranging in size from 34 to 140 kDa (Fig. 4A, lanes 2 and 4). None of these proteins binds to the domain B probe after transfer to nitrocellulose, however (Fig. 4B). Thus, the binding proteins either fail to renature or are present in very low amounts in this fraction. Another possibility is that they do not bind in 0.2 M KCl, which

was used for washing excess and non-specifically bound probe from the nitrocellulose.

Summary of the purification-The results of the purification are summarized in Table I. ABF-I (Fraction VIb) is enriched ~9,000-fold. The yield of ABF-I by this procedure is 225μg per 300g of cells, providing ample material for characterization of the protein and for production of antibodies.

Sequence specificity of binding in Fractions VIa, VIb and VIc-Initial experiments to characterize the binding sequence for each protein (ABF-I, Fraction VIa and VIb; and ABF-III, Fraction VIc) employed competition binding assays using three oligonucleotides. Two of the oligonucleotides correspond to element A and one to element B (see "Experimental Procedures"). Competition experiments were carried out so that the total amount of competitor, both oligonucleotide and neutral DNA, was equivalent in each reaction mixture. Neutral DNA used was either salmon sperm DNA or poly(dI:dC) and the results obtained were identical with each. As shown in Fig. 5, both Fraction VIa (Fig. 5B, lanes 4-7) and Fraction VIb (Fig. 5B, lanes 8-13) apparently recognize element B, since increasing amounts of element B oligonucleotide efficiently reduced complex formation, but the element A oligonucleotide competition did not (Fig. 5A, lanes 6-9; 14-17). Thus, Fractions VIa and b seem to be related.

When Fraction VIc was analyzed, two protein:DNA complexes were observed (Fig. 5A, lanes 2-4; Fig. 5B, lanes 1-3). Neither oligonucleotide competed for the proteins in the more rapidly migrating protein:DNA complex. Thus, Fraction VIc is apparently not recognizing either elements

A or B. Although the amount of the more slowly moving complex was reduced upon addition of the element A oligonucleotide, Fraction VIc' did not give a footprint at element A (data not shown). Fraction VIc behaved exactly as VIc' (data not shown).

DNase I footprint analysis of binding at ARS1-To further characterize these activities and locate their binding sites precisely within *ARS1*, standard DNase I footprint reactions were performed (Galas and Schmitz, 1978). The affinity purified 135kDa protein, Fraction VIb, protected a region between nucleotides 748 and 775 (Fig. 6A, lanes 8, 9, 14 and Fig. 6B, lane 6). DNase I protection at this site was eliminated when the element B oligonucleotide was added to the footprint reaction (data not shown), consistent with the competition binding experiments shown in Fig. 5. The DNase I protection at this site is similar to that for ABF-I seen by others (Shore *et al.*, 1987; Buchman *et al.*, 1988; Diffley and Stillman, 1988).

Fraction VIa gave a footprint identical to that seen with Faction VIb. Thus, binding activities in Fractions VIa and VIb were distinguishable by their chromatographic properties and the mobility of the protein-DNA complex observed in gel retardation assays (Figs. 3 and 4). However, these fractions protected exactly the same region of domain B from DNase I cleavage (Fig. 6).

Fig. 6 also shows that Fraction VIc gives a specific footprint on the *ARS1* fragment, and the protein responsible has therefore been designated ABF-III. Nucleotides 716 to 738, adjacent to the ABF-I site, were protected from DNase I cleavage by ABF-III, indicating that a novel *ARS1* binding

factor is present in this fraction. No protection at the adjacent ABF-I binding site was observed (Fig. 6) nor was the protection at the ABF-III site affected by addition of element B oligonucleotide. Partial protection was also observed around nucleotide 700 and nucleotide 880 in the *TRP1-ARS1* sequence. There are no sequence homologies between these weak sites and the strong ABF-III site, and it is not known whether protection at nucleotides 700 and 880 is due to weak binding of ABF-III or some other protein present in this impure fraction (see Fig. 5, lanes 2 and 4).

Since the ABF-I and ABF-III binding sites are within a 50bp region, we were interested in detecting possible interactions between the two factors. Mixing experiments (Fig. 6A, lanes 11 to 17) showed that ABF-I and ABF-III could bind simultaneously to their binding sites. The pattern of DNase I cleavage when both fractions were added was a composite of the individual footprints and the 10bp region between the binding sites remained DNase I sensitive (Fig. 6A, lanes 15 and 16). These data suggest that ABF-I and ABF-III can bind independently, though cooperative interactions cannot presently be ruled out. It should be noted that the specific binding observed here and the nature of the gel retardation complex distinguish ABF-III from ABF-II, a protein described previously (Diffley and Stillman, 1988). ABF-II binds *ARS1* DNA at multiple sites, and produces DNA:protein complexes that migrate much more rapidly in gel retardation assays than the ABF-III complex. Clarification of the relationship of ABF-III to ABF-II awaits further characterization of both proteins, however.

Fig. 6C summarizes the results of the DNase I footprint analysis in comparison with the published effects of domain B deletions on *ARS1* activity *in vivo*. Both ABF-I and ABF-III binding sites lie within the 3′ end of the *TRP1* coding region. Deletions removing most of the *TRP1* gene, including the ABF-III binding site, such as ΔH103, have been shown to have less than a 15% reduction in *ARS* activity when assayed in an *ARS*-CEN plasmid system (Srienc *et al.*, 1985). However, deletions that remove 50% of the ABF-I binding site reduce *ARS1* activity by 50%, suggesting that this protein plays a role in replication of *ARS1* containing plasmids. Deletions that remove the entire ABF-I site, however, do not lead to a significantly larger destabilization of the plasmid (Celniker *et al.*, 1984; Snyder *et al.*, 1986).

Binding to other ARSs-A detailed study of binding to the *ARS* adjacent to the silent mating type locus HMR, the HMRE-*ARS*, was undertaken to determine that ABF-I binding to element B was not an *ARS1*-specific phenomenon.

Standard footprinting procedures (Galas and Schmitz, 1978) were used to determine the recognition site in the HMR-E *ARS* and the results are shown in Fig. 7. The labeled fragments used in our studies contained both domains A and B. Fraction VIb protects the region between 281 to 256 bp, on the A-rich strand in the HMR-E fragment element B, exactly the same region as recognized by SBF-B or ABF-I (Shore *et al.*, 1987; Buchman *et al.*, 1988). The protected region on the other strand lies between 285 and 255bp (Fig. 7B).

DISCUSSION

We have purified a 135 kDa protein that binds sequence-specifically to *ARS1* and another *ARS*. DNase I footprinting shows that the protein protects an approximately 24bp region 82bp away from the *ARS* core sequence at *ARS1*. At the HMR-E *ARS* the protected region lies 66bp away from the *ARS* core. These are the same regions that are protected from DNase I cleavage by ABF-I or SBF-B (Shore *et al.*, 1987; Buchman *et al.*, 1988; Diffley and Stillman, 1988). In addition to the 135kDa protein, a chromatographically distinct, much less abundant species that also recognized the ABF-I site was detected during the purification. It is not yet clear whether it is an entirely different protein or a modified form of ABF-I.

ABF-I has an abundance almost equivalent to that of other replication proteins in the cell. Based on the amount of purified protein we obtained, we estimate there are about 700 molecules of the 135kDa ABF-I per diploid cell, comparable to the number of molecules of DNA polymerase I (Budd, Sitney and Campbell, unpublished).

Domain B has been shown to be necessary for efficient *ARS* function by deletion analysis (Bouton and Smith, 1986; Celniker *et al.*, 1984; Kearsey, 1984; Snyder *et al.*, 1986; Brand *et al.*, 1987) and transposon insertion mutation (Palzkill *et al.*, 1986). The importance of the ABF-I site itself is less clear. A 4bp deletion of the ABF-I site at *ARS1* decreases *ARS* activity only 10% (see Fig. 6C). A deletion covering one-half of the site reduces activity by 52%. Thus, ABF-I may not be essential for *ARS1* function, but does appear to be important. The ABF-I site has also been

shown to fall in a region important for *ARS* function at HMR-E (Brand *et al.*, 1987). The HMR-E *ARS* is of particular interest because either the *ARS* itself or sequences within it are essential for maintaining the repression of the MATα gene at the HMR locus. The functional elements within the HMR-E *ARS* have been defined by sequence comparison with other *ARS*s and by extensive mutagenesis (Abraham *et al.*, 1984; Shore *et al.*, 1987; Brand *et al.*, 1987). Like *ARS1*, the HMRE-*ARS* contains at least two elements, element A, a member of the core consensus (see Fig. 7) and the ABF-I site. As at *ARS1*, the ABF-I site lies about 60bp away from the core consensus, but at HMRE it is to the 5′ side of the T-rich strand, rather than to the 3′ side. It should also be added that the HMR-E *ARS* has some different properties from *ARS1*, making direct comparisons difficult (Kimmerly and Rine, 1987).

While binding at HMR-E and at *ARS1* is clearly due to the same protein, there is limited homology between the binding sequence found at the HMR-E *ARS* and ABF-I of *ARS1*. It is possible that the protein is recognizing some DNA structure that is shared by the different DNA sequences. A protein has been isolated from Crithidia fasciculata which recognizes bent DNA without recognizing a specific DNA sequence (Linial and Shlomai, 1987). It is interesting that Snyder *et al.* (1986) have provided evidence that the region at element B contains a bent structure that may be an important factor in binding. Bent DNA seems to be an important feature of many origins of replication (see references in Snyder *et al.*, 1986). Another possibility is that the recognition site is not a continuous sequence of base pairs. Twelve of 24 base pairs protected at the

HMR-E *ARS* are conserved at *ARS1*. The element B consensus proposed by Palzkill *et al.* (1986) is included in the region protected at *ARS1*, but is not noticeable at HMR-E. However, sequence analysis per se by us and others (Brand *et al.*, 1987; Palzkill *et al.*, 1986; Brand *et al.*, 1985; Buchman *et al.*, 1988) has not been able to define precisely the recognition signals in domain B. One advantage of having the purified protein in hand is that footprinting of the purified protein to additional *ARS*s will help define the consensus unambiguously, which is not possible with an impure fraction that may contain other DNA binding proteins such as ABF-III.

Since ABF-I binds to within 100bp at least five *ARS*s (Buchman *et al.*, 1988) and not just one, and since it binds to a functionally important region at HMRE and *ARS1*, it is probably important for the activity of these particular *ARS*s. Its availability in quantities made possible by the purification procedures described here paves the way for production of antibody to be used in a "reverse genetics" approach to verify this. For the moment, we can only speculate as to its precise role:

(1) Is it the initiator, the analogue of dnaA protein, λO protein or large T antigen? The problem with this idea is that one would have expected such a protein to recognize the essential core consensus of domain A, rather than the non-essential B.

(2) If not the initiatior protein, does ABF-I participate in transcriptional activation of some origins? Based on available data, this is perhaps the currently favored idea. Buchman *et al.* (1988) have shown that a second class of ABF-I sites is not associated with *ARS*s, but instead with transcriptional regulatory regions. Brand *et al.* (1987) have shown that

element B of the HMR-E *ARS* can activate transcription from the yeast CYC1 promoter. Transcriptional enhancers adjacent to eukaryotic viral origins of replication are important for DNA replication, though dispensable under some conditions. Sequences required for amplification of the chorion genes of *Drosophila* include a transcriptional promoter (for review see DePamphilis, 1988). Finally, some proteins may function in both transcription and replication (Jones *et al.*, 1987; DePamphilis, 1988).

(3) Another proposal is that the protein acts as a transcriptional terminator, and thus prevents transcription from interfering with replication (Snyder *et al.*, 1988). Termination of the *TRP1* mRNA occurs very near element B of *ARS1* (Fig. 4).

(4) Alternatively, one usually thinks of destabilization of *ARS* plasmids as being due to a failure to replicate, but instability could also be due to runaway replication. Negative regulatory factors have been invoked to account for the fact that each replicon is activated only once per cell cycle, an important factor in preventing the accumulation of chromosomal abnormalities. *ARS* plasmids, like chromosomal replicons, replicate only once per cell cycle (Fangman *et al.*, 1983). Thus, domain B proteins might prevent reinitiation at an origin that has already replicated.

The second *ARS* binding factor we have detected, ABF-III, is at this point less strongly implicated in *ARS* function than ABF-I, since the ABF-III recognition sequence can be deleted with little effect on replication *in vivo*. Furthermore, there appears to be no interaction between ABF-I and ABF-III. ABF-III may prove to be an ancillary protein for optimal

replication or a transcriptional termination factor at the end of the *TRP1* mRNA. Such studies await functional *in vitro* assays for *ARS* function.

Acknowledgments-We thank Daniel Derwin for technical assistance and Dr. Rati Verma for helpful discussions. This research was supported by the National Institutes of Health (GM-25508), March of Dimes (1-704), and American Cancer Society (NP-623E). K.S.S. was supported by the National Institutes of Health Cellular and Molecular Biology Training Program.

REFERENCES

Abraham, J., Nasmyth, K. A., Strathern, J. N., Klar, A. J. S., and Hicks, J. B. (1984) *J. Mol. Biol.* **176**, 307-331

Alberts, B., and Herrick, G. (1971) *Meth. Enzymol.* **21**, 198-217

Arndt-Jovin, D.J., Jovin, T.M., Böahr, W., Frischauf, A., and Marquardt, M. (1975) *Eur. J. Biochem.* **54**, 411-418

Bender, A., and Sprague, Jr., G. F. (1987) *Cell* **50**, 681-691

Bouton, A., and Smith, M. M. (1986) *Mol. Cell. Biol.* **6**, 2354-2363

Brand, A. H., Breeden, L., Abraham, J., Sternglanz, R., and Nasmyth, K. (1985) *Cell* 41, 41-48

Brand, A. H., Micklem, G., and Nasmyth, K. (1987) *Cell* **51**, 709-719

Brewer, B. J., and Fangman, W. L. (1987) *Cell* **51**, 463-471

Broach, J. R., Li, Y. -Y., Feldman, J., Jayaram, M., Abraham, J., Nasmyth, K. A. and Hicks, J. B. (1983) *Cold Spring Harbor Symp. Quant. Biol.* **47**, 1165-1173

Buchman, A. R., Kimmerly, W. J., Rine, J., and Kornberg, R. D. (1988) *Mol. Cell. Biol.* **8**, 210-225

Campbell, J. L., Budd, M., Gordon, C., Jong, A. Y. S., Sweder, K., Oehm, A., and Gilbert, M. (1986) in *Mechanisms of DNA Replication and Recombination* (Kelly, T., McMacken, R., eds) Vol. 47 pp. 265-287, UCLA Symp. on Molecular and Cellular Biology, N.Y. Series, Alan R. Liss, Inc., New York

Celniker, S. E., and Campbell, J. L. (1982) *Cell* **31**, 201-213

Celniker, S. E., Sweder, K., Srienc, F., Bailey, J. E., and Campbell, J. L. (1984) Mol. *Cell. Biol.* 4, 2455-2466

Chan, C. S. M., and Tye, B. -K. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 6329-6333

DePamphilis, M. L. (1988) *Cell* **52**, 635-638.

Diffley, J. F. X., and Stillman, B. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2120-2124

Fangman, W. L., Hice, R. H., and Chlebowicz-Sledziewska, E. (1983) *Cell* **32**, 831-838

Fried, M., and Crothers, D. M. (1981) *Nucl. Acids Res.* **9**, 6505-6525

Fuller, R. S., Funnell, B. E., and Kornberg, A. (1984) *Cell* **38**, 889-900

Galas, D. J., and Schmitz, A. (1978) *Nucl. Acids. Res* **5**, 3157-3170

Garner, M. M., and Revzin, A. (1981) *Nucl. Acids Res.* **9**, 3047-3060

Huberman, J. A., Spotila, L. D., Nawotka, K. A., El-Assouli, S. M., and Davis, L. R. (1987) *Cell* **51**, 473-481

Jacob, F., Brenner, S., and Cuzin, F. (1963) *Cold Spring Harbor Symp. Quant. Biol.* 28, 329-348

Johnson, L. M., Snyder, M., Chang, L. M. S., Davis, R. W., and Campbell., J. L. (1985) *Cell* **43**, 369-377

Jones, K. A., Kadonaga, J. T., Rosenfield, P. J., Kelly, T. J., and Tjian, R. (1987) *Cell* **48**, 79-89

Jong, A. Y. -S., Aebersold, R., and Campbell, J. L. (1985) *J. Biol. Chem.* **260**, 16367- 16374

Kadonaga, J. T., and Tjian, R. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 5889-5893

Kearsey, S. (1984) *Cell* **37**, 299-307

Kimmerly, W. J., and Rine, J. (1987) *Mol. Cell. Biol.* **7**, 4225-4237

Koshland, D., Kent, J. C., and Hartwell, L. H. (1985) *Cell* **40**, 393-403

Li, J. J., Peden, K. W. C., Dixon, R. A. F., and Kelly, T. (1986) *Mol. Cell, Biol.* **6** 1117-1128.

Li, J. J., and Kelly, T. J. (1985) *Mol. Cell. Biol.* **5**, 1238-1246

Linial, M., and Shlomai, J. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 8205-8209

Marunouchi, T., Matsumoto, Y. -I., Hosoya, H., and Okabayashi, K. (1987) *Mol. Gen. Genet.* **206**, 60-65

Maxam, A. M., and Gilbert, W. (1980) *Enzymol.* **65**, 499-560

Miskimins, W. K., Roberts, M. P., McClelland, A., and Ruddle, F. H. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 6741-6744

Montiel, J. F., Norbury, C. J., Tuite, M. F., Dobson, M. J., Mills, J. S., Kingsman, A. J., and Kingsman, S. M. (1984) *Nucl. Acids Res.* **12**, 1049-1067

Newlon, C. S., Devenish, R. J., Suci, P. A., and Roffis, C. J. (1981) in *The Initiation of DNA Replication* (Ray, D.S., ed) Academic Press, New York

Newlon, C. S., and Burke, W. (1980) in *Mechanistic Studies of DNA Replication and Genetic Recombination* (Alberts, B., Fox, C.F., ed) ICN-UCLA Symposia on Molecular and Cellular Biology, 19, Academic Press, New York

Palzkill, T. G., and Newlon, C. S. (1988) *Cell* **53**, 441-450

Palzkill, T. G., Oliver, S. G ., and Newlon, C. S. (1986) *Nucl. Acids Res.* **14**, 6247-6264

Petes, T. D., and Williamson, D. H. (1975) *Cell* **4**, 249-253

Rivin, C. J., and Fangman, W. L. (1980) *J. Cell. Biol.* **85**, 108-115

Saffer, L. D., and Miller, Jr., O. L. (1986) *Mol. Cell. Biol.* **6**, 1148-1157

Shore, D., Stillman, D. J., Brand, A. H., and Nasmyth, K. A. (1987) *EMBO J.* **6**, 461-467

Snyder, M., Buchman, A. R., and Davis, R. W. (1986) *Nature* **324**, 87-89

Snyder, M., Sapolsky, R. J., and Davis, R. W. (1988) *Mol. Cell. Biol.* **8**, 2184-2194

Srienc, F., Bailey, J. E., and Campbell, J. L. (1985) *Mol. Cell. Biol.* **5**, 1676-1684

Stinchcomb, D. T., Mann, C., Selker, E., and Davis, R. W. (1981) in *The Initiation Of DNA Replication* (Ray, D.S., ed) p. 473-488, ICN-UCLA Symposia on Molecular and Cellular Biology, 22, Academic Press, New York

Struhl, K., Stinchcomb, D. T., Scherer, S., and Davis, R. W. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 1035-1039

Tschumper, G., and Carbon, J. (1980) *Gene* **10**, 157-166

Weber, K., and Osborn, M. (1969) *J. Biol. Chem.* **244**, 4406-4412

Wiederrecht, G., Shuey, D. J., Kibbe, W. A., and Parker, C. S. (1987) *Cell* 48, 507-515

Wold, M. S., Li, J. J., and Kelly, T. J. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 3643-3647

Yasuda, S., and Hirota, Y. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5458-5462

Zyskind, J. W., Cleary, J. M., Brusilow, W. S. A., Harding, N. E., and Smith, D. W. (1983) *Proc. Natl. Acad. Sci. USA* 80, 1164-1168

Table 1. Purification of ABF-I and ABF-III from 300 g of yeast cells.

Fraction	Step	Protein mg	Protein mg/ml	Specific Activity ng fragment bound / µg Protein	Recovery of Activity (%)
I	Crude lysate	30,072	35.8	N.D.	--
II	DNA free extract	15,353	11.5	0.0008	100
III	Native DNA Cellulose	61	0.39	0.209	100
IV	Phosphocellulose	13	0.18	1.03	100
V	Mono Q				
Va	0.27 M KCl	0.84	0.16	3.47	18
Vb	0.33 M KCl	1.48	0.19	8.96	82
VI	Oligonucleotide Affinity				
VIa (ABF-I)	0.7 M KCl	0.019		6.64	1.05
VIb (ABF-I)	0.7 M KCl	0.225		7.11	13.4
VIc (ABF-III)	0.2 M KCl	0.598		0.441	2.20

Activity was measured by densitometric tracing of bands in protein:DNA complexes in standard gel retardation assays. For accuracy, three different concentrations of protein which retarded approximately 20, 40, and 60% of the labeled fragment, respectively, were used for each determination. The "ABF" designation in Fraction VI parentheses refers to the site recognized.

FIGURE LEGENDS

Fig. 1. **Physical map of *ARS1* and Sequence of Core Element.** The extent of domains A, B and C was originally derived from Celniker *et al.* (1984) and refined from numerous references in the text. Coincident with the 3′ end of the *TRP1* mRNA, indicated by the upper arrow, are a bent DNA region and the ABF-I binding site as reported in Snyder *et al.* (1986) and Buchman *et al.* (1988). Domain A contains the 11bp core consensus element shown (Broach *et al.*, 1983). Close matches to the core consensus are indicated with solid boxes above the line with the number of matches given below the line. Small arrows above the matches indicate orientation of the sequence relative to the core.

Fig. 2. **Analysis of *ARS* binding activities eluting from Mono Q.** Fractions eluting between 0.3 and 0.4 M KCl were assayed by gel retardation as described in "Experimental Procedures" using the EcoRI-Hind III fragment of YRp09ΔH103 as the probe. Each binding reaction contained 1μl of column fraction in a 50μl reaction volume with 2ng of labeled probe. All reactions contained 1μg of salmon sperm DNA. Free DNA is the fastest migrating discrete band. Only this band appears in a minus protein control (see Fig. 3, lanes 4, 10, 16).

Fig. 3. **Oligonucleotide affinity chromatography.** A) Fraction Va was chromatographed on a 1 ml element B oligonucleotide affinity

column. Column fractions were assayed by gel retardation by incubating 2μl of each fraction with 0.8 ng labeled EcoRI-HindIII ΔH103 fragment as described in "Experimental Procedures." A slowly migrating complex, ABF-III, elutes at low salt while a faster migrating ladder of complexes elutes at high salt. B) Fraction Vb, which gives rise to the more slowly migrating complex, was chromatographed on a 2ml element B oligonucleotide affinity column as in A). Gel retardation assays with 1μl of each fraction and the same probe as in A) yielded two chromatographically distinguishable, slowly migrating complexes. The first elutes at the same low salt as ABF-III in A). The second slowly migrating complex elutes at high salt. All reactions contained 1μg of salmon sperm DNA as neutral competitor. F, flow through; W, wash; S, sample loaded; B, blank (no protein).

Fig. 4. **Analysis of *ARS* binding proteins by gel electrophoresis and "Southwestern" blotting.** A) Element B oligonucleotide affinity column fractions were analyzed on a 7.5% SDS-polyacrylamide gel. Protein bands were visualized by silver staining. Lane 1, Fraction VIb (ABF-I); lane 2, Fraction VIc (ABF-III from Fraction Vb); lane 3, Fraction VIa (ABF-I); lane 4, Fraction VIc, ABF-III from Fraction Va. B) "Southwestern blotting" was performed as in "Experimental Procedures" using the BglII-HindIII fragment of YRp7 as probe. Lane 1, Fraction VIb (ABF-I) (4μg); lane 2,

Fraction VIc (ABF-III) (21.5μg); lane 3, Fraction VIa (ABF-I) (1.08μg); lane 4, Fraction VIa (ABF-III) (36.0μg). Inset, Southwestern blot of Fraction VIb (4μg) that had been subjected to electrophoresis on a separate gel under conditions allowing resolution of the binding activity into two bands (i.e., the gel was run 30% further).

Fig. 5. **Oligonucleotide competition binding assays.** Binding experiments were carried out as described in "Experimental Procedures." A) Oligonucleotide affinity column protein fractions were incubated with 0.8ng labeled EcoRI-HindIII fragment from ΔH103 and salmon sperm DNA or element A oligonucleotide and salmon sperm DNA. Total DNA concentration was kept constant at 1μg. Lane 1, no protein; lanes 2-5, 200ng Fraction VIc′ (from Vb) with 0ng, 30ng, 100ng, and 300ng element A oligonucleotide; lanes 6-9, 18ng Fraction VIa with amounts of oligonucleotide as above; lanes 10-13, 179ng of Fraction VIc (from Vb); lanes 14-17, 26.4ng of Fraction VIb. B) Oligonucleotide affinity column protein fractions were incubated exactly as above except element B oligonucleotide was used instead of element A oligonucleotide. Lanes 1-3, 200ng of Fraction VIc′ (from Vb) with 0ng, 300ng, and 500ng of element B oligonucleotide; lanes 4-7, 18ng of fraction VIa with 0ng, 30ng, 100ng, and 300ng of element B oligonucleotide; lanes 8-10, 251ng of Fraction VIc′ (from Vb) with 0ng, 300ng, and 500ng of element B oligonucleotide; lanes 11-13, 26.4ng of

Fraction VIb with 0ng, 300ng, and 500ng of element B oligonucleotide.

Fig. 6. **DNase I footprint of affinity-purified fractions on *ARS1*.** The 237bp HindIII-BglII fragment of *ARS1* was 5′ end labeled at either the HindIII (A) or BglII site (B). Each reaction contained 6ng of labeled DNA and 1μg of non-specific salmon sperm DNA in a 30μl reaction volume. Protein fractions were allowed to bind for 10 min at 4°C, followed by addition of 10ng of DNase I and digestion for 1 min at room temperature. (A) Lanes 1, 4, 7, 10, 11, 17 contain no protein; Lanes 2, 3: Fraction VIa -- 0.05μg, 0.18μμg of protein, respectively; Lanes 5, 6: Fraction VIc -- 1.1μg, 3.6μg of protein; Lanes 8, 9: Fraction VIb -- 0.2-μg, 0.66μg of protein; Lane 12: Fraction VIa -- 0.09μg of protein; Lane 13: Fraction VIc -- 1.8μg of protein; Lane 14: Fraction VIb -- 0.33μg of protein; Lane 15: Fraction VIc + VIa -- 1.8μg and 0.09μg of protein, respectively; Lane 16: Fraction VIc + VIb -- 1.8μg and 0.33μg of protein, respectively. (B) Lanes 1, 4, 7 contain no protein; Lane 2: Fraction VIc -- 6μg of protein; Lane 3: Fraction VIa -- 0.18μg of protein; Lane 5: Fraction VIc -- 3.6μg of protein: Lane 6: Fraction VIb -- 0.66μg of protein. Marker lanes (G or Y) contain Maxam-Gilbert G or G+A specific sequencing ladders for either fragment. (C) Location of ABF-I and ABF-III binding sites within domain B of *ARS1*. The brackets on each strand indicate the extent of DNase I protection in the presence of each binding factor.

Numbering of the *TRP1-ARS1* sequence was based on Tschumper and Carbon (1980). The TAG at position 775 codes for the predicted termination codon of the *TRP1* gene. Shown below the sequence is a summary of previously published results (Srienc *et al.*, 1985; Snyder *et al.*, 1986; Diffley and Stillman, 1988; Celniker *et al.*, 1984) on the effects of deletion mutagenesis on *ARS* activity. The arrows represent the end points of 5′ deletions into domain B where domain A was left intact. A ⊕ indicates that the mutant *ARS1* cloned in a CEN plasmid had a reported plasmid loss per generation of less than 15%; ⊖ indicates that the mutant *ARS1*-CEN plasmid loss per generation was greater than 45%.

Fig. 7. **DNase I footprint of the *HMR-E ARS*.** The 450bp XhoI-XbaI fragment of HMR-E *ARS* deletion mutant pJA82.6Δ173 (Abraham *et al.*, 1984) was 5′ end labeled at either the XbaI (A) or XhoI site (B). The reactions were carried out as described in Figure 6 except 500ng of salmon sperm DNA was used. (A) Lane 1 contains no protein or DNase I; Lanes 2-6 are DNase I digests; Lanes 2 and 6 contain no protein; Lanes 3-5: Fraction VIb -- 56ng, 168ng, and 560ng of protein, respectively. Lanes marked G and A+G contain products of G reactions or A+G reactions (Maxam and Gilbert, 1980). The protected region extends from 257 to 280 on this strand. (B) Lane 1 contains no protein or DNase I; Lanes 2-6 are DNase I digests; Lanes 3-5: Fraction VIb -- 53ng, 105ng and

210ng of protein, respectively. The protected region extends from 256 to 284 on this strand.

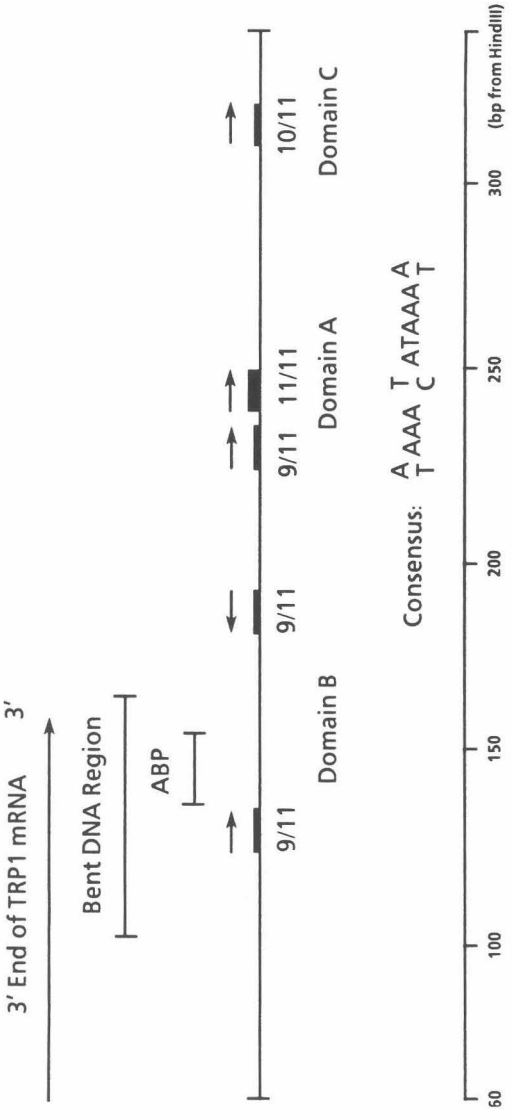

3' End of TRP1 mRNA
3'
Bent DNA Region
ABP
9/11
9/11
9/11
11/11
10/11
Domain B
Domain A
Domain C
Consensus: A/T AAA T/C ATAAA A/T
60
100
150
200
250
300
(bp from HindIII)

Fig. 1.

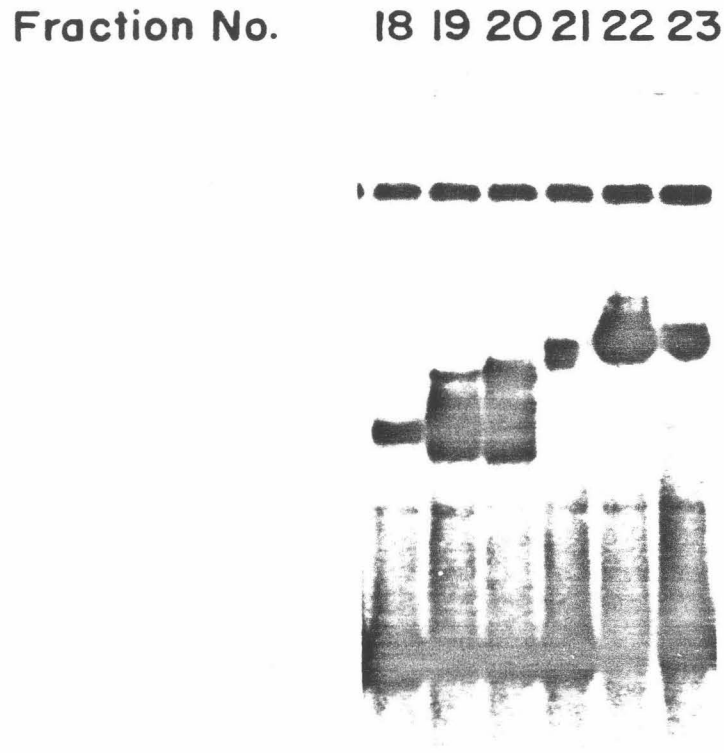


Fig. 2.

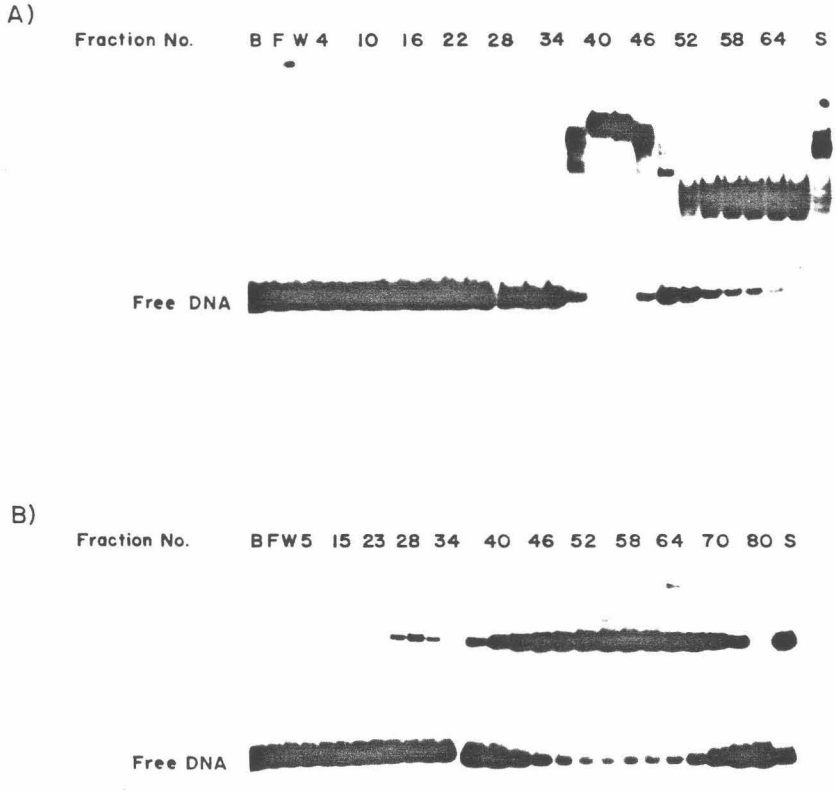

A)
Fraction No.
B F W 4 10 16 22 28 34 40 46 52 58 64 S
Free DNA
B)
Fraction No.
BFW5 15 23 28 34 40 46 52 58 64 70 80 S
Free DNA

Fig. 3.

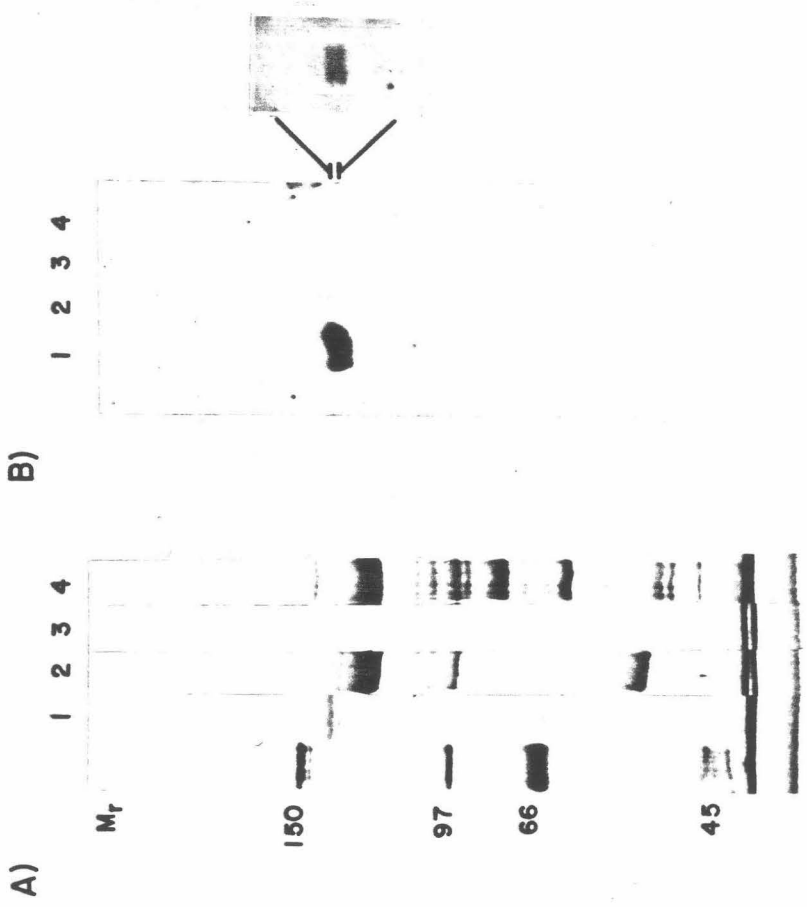
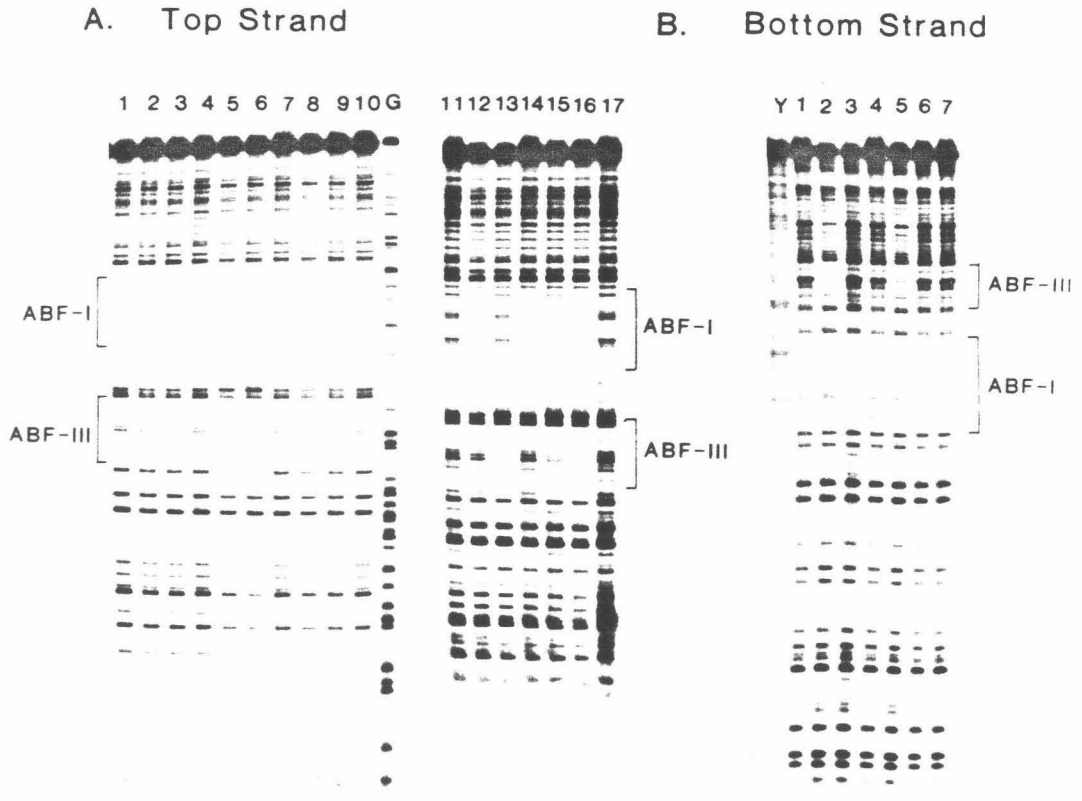


Fig. 4.


A)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
Free DNA
B)
1 2 3 4 5 6 7 8 9 10 11 12 13
Free DNA

Fig. 5.


A. Top Strand
B. Bottom Strand
1 2 3 4 5 6 7 8 9 10 G
11 12 13 14 15 16 17
Y 1 2 3 4 5 6 7
ABF-I
ABF-III
ABF-I
ABF-III
ABF-III
ABF-I

C.

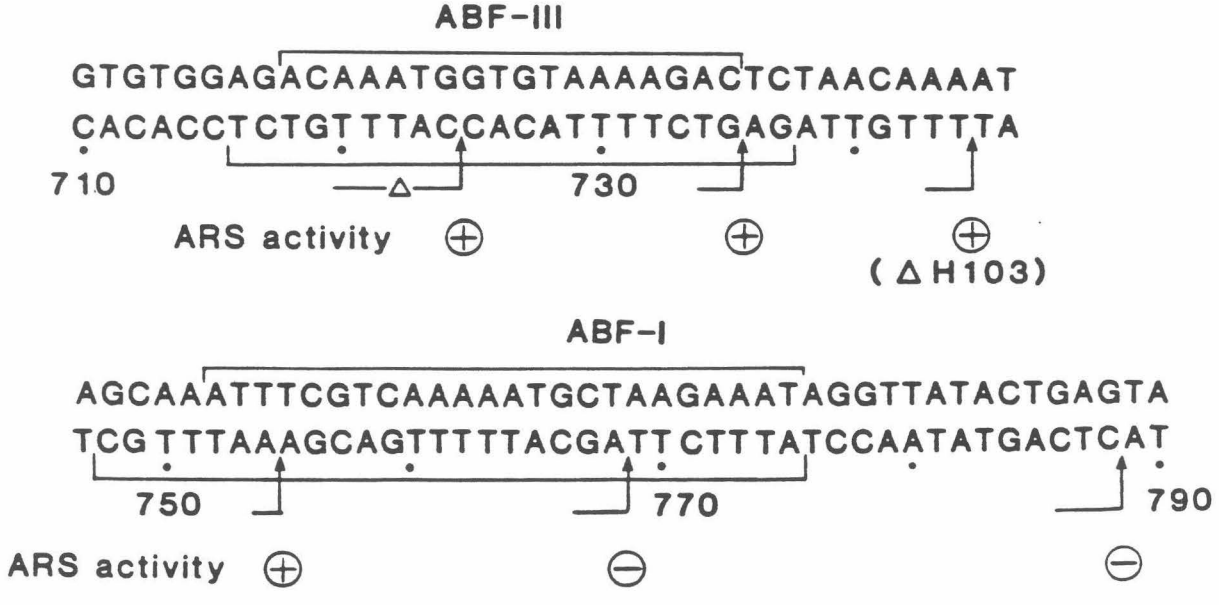

ABF-III
GTGTGGAGACAAATGGTGTAAAAGACTCTAACAAAAT
CACACCTCTGTTTACCACATTTTCTGAGATTGTTTTA
710
730
ARS activity
(ΔH103)
ABF-I
AGCAAATTTCGTCAAAAATGCTAAGAAATAGGTTATACTGAGTA
TCGTTTAAAGCAGTTTTTACGATTCTTTATCCAATATGACTCAT
750
770
790
ARS activity

Fig. 6.

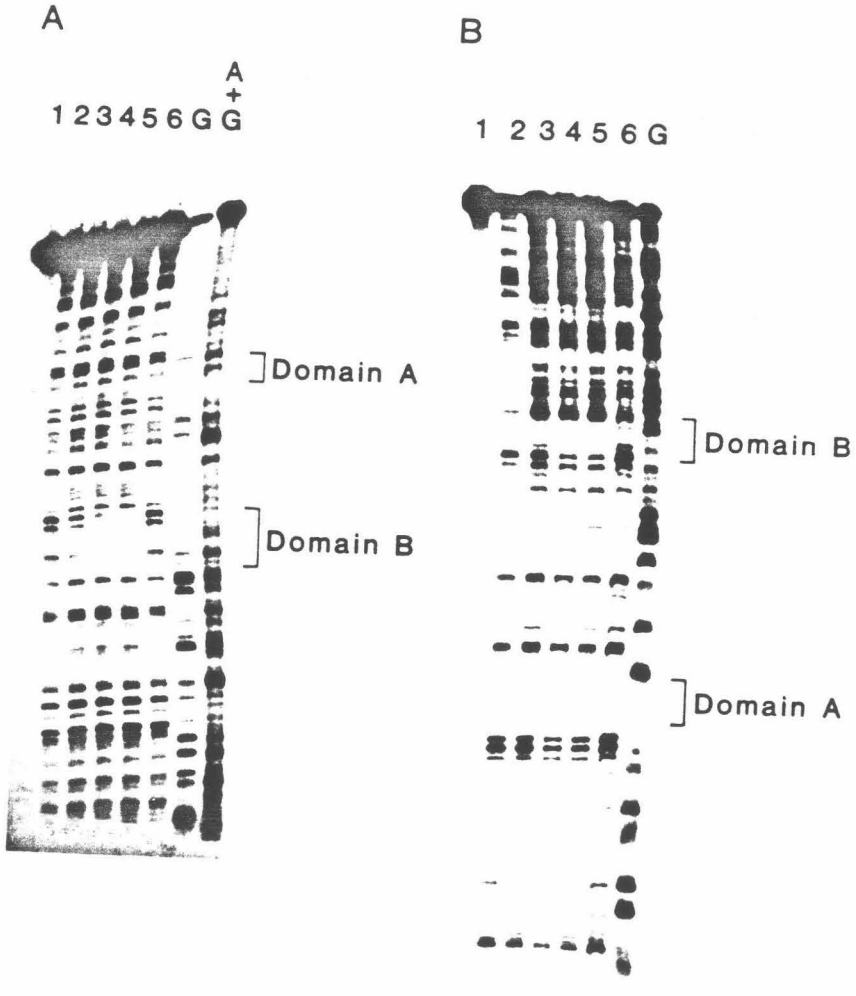


Fig. 7.

Supplementary Material to:

Purification and Characterization of Proteins That Bind To Yeast *ARSs*

EXPERIMENTAL PROCEDURES

Materials-Acrylamide was from Serva Feinbiochemica. T4 DNA ligase, T4 polynucleotide kinase, calf thymus DNA and restriction endonucleases were from United States Biochem. SDS was from IBI. Poly(dI:dC) was from Boehringer. Salmon sperm DNA was from Sigma. The Mono Q HR 5/5 column, Sepharose, and all nucleotides were from Pharmacia LKB Biotechnology, Inc. [γ-^{32}P]ATP was from Amersham. P11 (phosphocellulose) was from Whatman. Polyethylene glycol 8000 was from Carbowax. Yeast extract and Bacto-peptone were from Difco Laboratories. Cerelose (Dextrose) was from Corn Products Corp.

Strains-Saccharomyces cerevisiae PEP4D (*MATa/α, his1/+, trp1/+, pep4-3/pep4-3, prc1-126/prc1-126, prb1-1122/prb1-1122, can1/can1*) is a protease-deficient strain that was used for isolation of *ARS*-binding proteins (Johnson *et al.*, 1985).

Growth of cells and preparation of DNA-free extracts-Initial steps in the purification of yeast *ARS*-binding proteins were based on the methods of Alberts and Herrick (1971), with modifications of Jong *et al.* (1985). Yeast cells (PEP4D) were grown in a 350 ℓ fermenter to mid-log phase in YPD medium [1% yeast extract (w/v), 2% peptone (w/v), and 2% cerelose (w/v)]

at 30°C. The cells were harvested in a Sharples centrifuge and the cell pellets were frozen in liquid nitrogen and stored at -70°C.

For a standard preparation, 300g of cells were resuspended in 800ml of buffer A (20mM HEPES, pH7.4, 10% glycerol, 1mM 2-mercaptoethanol, 1mM EDTA, 1mM phenylmethyl sulfonyl fluoride) containing 1M KCl and protease inhibitors (1mM EGTA, 2μg/ml pepstatin A, 1μg/ml soybean trypsin inhibitor, 1mM benzamidine, 1μg/ml leupeptin and 50mM diisopropyl fluorophosphate). All operations were carried out at 4°C. Cells were lysed by passing them 3-5 times through a Dyno-Mill homogenizer (Glen Mills, Inc.). The lysate was centrifuged at 12,000 rpm for 45 min in a Sorvall GSA rotor and then filtered through four layers of cheese cloth. The resulting solution was designated Fraction I. Polyethylene glycol 8000 (30%, w/v) in 2M KCl was added to Fraction I (840 ml) to a final concentration of 6% (w/v). After gentle stirring for 45 min, the nucleic acid-containing precipitate was removed by centrifugation at 12,000 rpm in a Sorvall GSA rotor for 25 min. The supernatant was dialyzed for 17 hr against 16 ℓ buffer A containing 100mM KCl and the protease inhibitors described above. Precipitated material was removed by centrifugation for 10 min at 8,000 rpm, the supernatant was designated Fraction II.

Chromatographic procedures-Fraction II (1335 ml) was loaded onto a native DNA-cellulose column (24 cm x 12.5 cm^2) prepared according to Alberts and Herrick (1971). The column was washed with 3 ℓ of buffer A containing 0.1M KCl. Proteins that remained bound were eluted with a 1.5 ℓ linear gradient from 0.1M to 1.0M KCl in buffer A. All fractions were assayed for *ARS1* specific DNA binding activity in the presence of 5μg of

sonically irradiated salmon sperm DNA or poly(dI:dC). All fractions were also assayed for DNA polymerase activity. Fractions were pooled so as to exclude as much of the polymerase activity as possible. Fractions eluting between 0.3 and 0.4M KCl, which contained the *ARS*-binding proteins, were pooled and dialyzed 7 hr against buffer A containing 0.1M KCl. The dialyzed sample was designated Fraction III.

Fraction III (156 ml) was loaded onto a phosphocellulose column (7.85 cm x 5.73 cm^2) previously equilibrated with buffer A containing 0.1M KCl. The column was washed with 270ml of the same buffer and the proteins were eluted with a 450ml linear gradient from 0.1M to 1.0M KCl in buffer A. Again each fraction was assayed for *ARS1* specific DNA binding activity in the presence of 5µg salmon sperm DNA and each fraction was assayed for DNA polymerase activity. Fractions eluting between 0.37-0.52M KCl were pooled and designated IV. Fractions were pooled to exclude any residual DNA polymerase. The pooled fractions were dialyzed 5 hr against 20mM Tris·HCl, pH8.0, 50mM KCl, 5% glycerol, and 1mM β-mercaptoethanol (buffer B).

Fraction IV (74ml) was loaded onto an FPLC Mono Q HR 5/5 column previously equilibrated with buffer B. Proteins were eluted with a 50ml linear gradient from 50mM to 1.0M KCl in buffer B. Assays contained 1µg salmon sperm DNA competitor. Fractions eluting between 0.24-0.29M and 0.29-0.36 KCl were pooled separately and labeled Va and Vb, respectively. The pools were dialyzed 5 hr against buffer A containing 0.1M KCl.

Dialyzed Fraction Va (5.3ml) was loaded onto a 1 ml oligonucleotide-Sepharose column prepared as described below and washed with 20ml of

buffer A containing 0.1M KCl. Protein was eluted with a 10ml gradient from 0.1M to 0.9M KCl in buffer A. Binding activities eluting between 0.15-0.33M and 0.55-0.86M KCl were pooled separately and designated fractions VIc and VIa, respectively.

Fraction Vb (7.9ml) was loaded onto a 2ml oligonucleotide-Sepharose column. The column was washed with 40ml of buffer A containing 0.1M KCl. Protein was eluted with a 20ml gradient from 0.1M to 0.9M KCl in buffer A. Proteins eluting between 0.2-0.31M and 0.43-0.79M were pooled separately and labeled fractions VIc and VIb, respectively.

Fractions VIa, b and c were dialyzed against buffer A containing 50% glycerol and stored at -20°C.

Preparation of DNA for coupling to Sepharose-Chemically synthesized complementary oligonucleotides were annealed, 5′-phosphorylated, and ligated by the method of Kadonaga and Tjian (1986). Oligomers of the original oligonucleotide consisted of a mixture of 3-mers to 20-mers, with greater than 80% in the 20-mer length range.

Coupling of oligonucleotides to Sepharose-The ligated oligonucleotides were coupled to Sepharose CL-6B by the method of Arndt-Jovin *et al.* (1975). For the column described in Fig. 3, 3.5mg of double stranded oligonucleotide and 7ml (settled volume) activated Sepharose CL-6B were used.

Renaturation of binding activity from SDS gels-The binding species in each protein preparation were determined by the protein-blotting procedure of Miskimins *et al.* (1985) with slight modifications. Protein samples were mixed with one-fourth volume of 4x sample buffer (Weber

and Osborn, 1969). Samples were not boiled before electrophoresis. Proteins were electrophoresed in 7.5% polyacrylamide (82:1 acrylamide:bisacrylamide) gels containing SDS at 4°C and then transferred to nitrocellulose. Transfer was carried out at 4°C in 25mM Tris base/190mM glycine for 10 hr at 100 mA.

Nitrocellulose strips were incubated with 2.5% (wt/vol) non-fat dry milk in binding buffer (1mM Tris, pH 7.4/50mM KCl/5% glycerol/0.05% Nonidet P-40) for 1 hr at 4°C, washed with binding buffer and incubated 8 hr in binding buffer with 0.25% non-fat dry milk containing ^{32}P-labeled *ARS* DNA and 25μg/ml salmon sperm DNA. Free DNA fragments were removed by washing two times for 5 min in 0.25% (wt/vol) non-fat dry milk in binding buffer and once for 4.5 min in 0.25% (wt/vol) non-fat dry milk/200 mM KCl in binding buffer at 4°C. Filters were wrapped in cellophane and radioactivity bound to proteins on the nitrocellulose was determined by autoradiography using Kodak XAR-5 film.

Gel retardation assay-DNA-binding activity was monitored by altered electrophoretic mobility of the protein-DNA complex (Fried and Crothers, 1981; Garner and Revzin, 1981). Protein samples were incubated with 0.6-2.0ng of [^{32}P] DNA in binding buffer (10mM Tris, pH7.4/50mM KCl/5% glycerol/0.05% Nonidet P-40), on ice or at room temperature. All reactions contained at least 500ng sonically irradiated salmon sperm DNA except where indicated.

Preparation of labeled fragments-Yeast *ARS1* DNA and yeast *HMR-E ARS* DNA were used for both footprinting and gel retardation assays. All DNAs were labeled using T4 polynucleotide kinase. The *ARS1*

deletion mutant YRp09ΔH103 (Celniker *et al.*, 1984) was labeled at either the EcoRI or HindIII site and then cleaved with HindIII or EcoRI , respectively. This generated a ~234 bp fragment, containing domains A, B, and 117 base pairs of domain C.

ARS1-derived plasmid YRp09ΔH200 (Celniker *et al.*, 1984) was labeled at the EcoRI or HindIII site, then cleaved with HindIII or EcoRI to generate a 163bp fragment. This fragment contains domain A, domain C, and most of domain B, but is lacking element B as defined by Palzkill and Newlon (1988). YRp09 ΔH200 replicates very inefficiently (Celniker *et al.*, 1984).

The *ARS1* plasmid YRp7 (Struhl *et al.*, 1979) was labeled at the BglII or HindIII site and then cleaved with HindIII or BglII. The fragment generated was 237bp long and contained domain B only.

HMR-E plasmid pJA82.6Δ173 (Abraham *et al.*, 1984) was labeled at XhoI or XbaI restriction sites and cleaved with XbaI or XhoI, respectively, to generate a 450bp fragment. The fragment contains domains A, E, and B. Element E is defined in Shore *et al.* (1987).

Oligonucleotides-The element A oligonucleotides were

GAATTCCAGATTTTATGTTTAGATC
GTCTAAAATACAAATCTAGCTTAAG

and

TTTTACAGATTTTATGTTTAGATCTTTTATGCTTG
GAACAAAATGTCTAAAATACAAATCTAGAAAATAC

The element B oligonucleotide (ABF-I site) was

GAATTCATTTCTTAGCATTTTTGACGAAATTTG
TAAAGAATCGTAAAAACTGCTTTAAACCTTAAG.

This element B oligonucleotide was used for the affinity column.

CHAPTER 3

Isolation of the Gene Encoding *ARS* Binding Factor I (ABF-I)

ABSTRACT

The *ARS* binding factor I (ABF-I) has been implicated in transcriptional repression and DNA replication. Using rabbit polyclonal antisera and murine monoclonal antibodies against ABF-I we have screened a yeast expression library. Four EcoRI restriction fragments were isolated that encoded proteins which were recognized by both polyclonal and monoclonal antibodies.

INTRODUCTION

Initiation of eukaryotic DNA replication occurs at numerous origins within a defined portion of the cell cycle called S phase. Each origin is used only once per S phase, implying that there must be some kind of temporal control of initiation of DNA replication. The mechanism of control is not known. However, isolation of those proteins that interact with origins of replication should help elucidate the regulatory pathway for temporal control of initiation.

The sequences that enable extrachromosomal maintenance of any colinear DNA in yeast are called autonomously replicating sequences (*ARS*s). *ARS*s have been shown to be used as replication origins on plasmids *in vitro* (Celniker and Campbell, 1982) and *in vivo* (Brewer and Fangman, 1987; Huberman *et al.*, 1987). Deletion analysis (Celniker *et al.*, 1984; Kearsey, 1984; Bouton and Smith, 1986) and transposon insertion mutagenesis (Palzkill, 1986) revealed three functional domains, A, B, and C, which allow for full replication. Domain C is the region where replication bubbles form *in vitro* (Celniker and Campbell, 1982). Domain C can be deleted with minimal effect on replication. Domain A is an 11 to 19bp sequence containing an 11bp sequence, element A, that is conserved at all *ARS*s. Domain A is required and sufficient for *ARS* function on centromeric plasmids; however, plasmids carrying only the domain A *ARS* sequence replicate inefficiently. Domain B extends 50 to 100bp 3′ to domain A on the T-rich strand. Domain B contains a sequence, element B, about 80bp from element A which is recognized by a 135kDa protein, *ARS* binding factor I (ABF-I) (Sweder *et al.*, in press; Diffley and Stillman,

1988). Domain B is crucial for efficient replication but is not sufficient for *ARS* activity.

ARS Binding Factor I (ABF-I) is the first example of a protein that interacts sequence-specifically with a eukaryotic origin of replication (Sweder *et al.*, 1988; Diffley and Stillman, 1988). The protein has an apparent molecular weight of 135kDa and is present at ~700 copies per diploid cell. It is a nuclear protein based on *in situ* antibody labelling (Michael Clark, personal communication). The degree of staining with antibody is in good agreement with the copy number estimate from protein yield when compared to other low abundance proteins such as tRNA ligase. The role of ABF-I in replication is not known. The sequence recognized by ABF-I is not present at all *ARS*s and there is little homology between the ABF-I binding sites at different *ARS*s. Additionally, ABF-I binding sites are found not associated with *ARS*s in the yeast genome (Buchman *et al.*, 1988).

The possibility exists that ABF-I is a transcription factor that is involved in replication. ABF-I binding sites are present in the HMR-E silencer sequences required for transcriptional repression of the silent mating locus HMR. HMR-E has an associated *ARS* activity that is required for repression of transcription (Miller and Nasmyth, 1984). ABF-I may be involved in transcription termination. By placing *ARS1* downstream from a strong yeast promoter, either *GAL1* or *GAL10*, Snyder *et al.* (1988) were able to study the effect of transcription on *ARS* activity. Induction of the *GAL* promoter resulted in a drastic reduction in mitotic stability of otherwise stable artificial yeast chromosomes. Introduction of

a transcription termination region from the 3' end of the CAT gene between the promoter and *ARS1* restored replication of the artificial chromosomes.

Alternatively, ABF-I binding to its recognition site could act as a transcriptional activator. Transcriptional activation of replication by RNA polymerase has already been demonstrated in *E. coli* (Baker and Kornberg, in press). The promoter and enhancer sequences flanking the SV40 origin of replication have a dramatic effect on the efficiency of replication in competent cells (Wold *et al.*, 1987; Dean *et al.*, 1987). Transcriptional activation by elements within yeast *ARS*s has been investigated for sequences from the HMR-E *ARS*. The ABF-I binding site was cloned into a plasmid carrying a *cyc1::lacZ* fusion gene lacking its own *UAS* (Brand *et al.*, 1987). Transcriptional activation by the ABF-I site was quantified by a β-galactosidase assay. The presence of one or two ABF-I sites increased transcription levels by 4.4 and 6.4 fold, respectively.

To determine the role of ABF-I in yeast DNA replication we have cloned the gene encoding ABF-I. This was done by screening a λgt11 yeast expression library with rabbit polyclonal antibodies to ABF-I. Positives were rescreened with mouse monoclonal antibodies to ABF-I to confirm the presence of the ABF-I epitope in the λgt11 clones. With the gene in hand, the chromosomal copy of the ABF-I gene can be deleted and the phenotype of the resulting spores determined. If ABF-I is required for transcriptional termination or transcriptional activation of DNA replication, it is likely to be an essential gene.

MATERIALS AND METHODS

Strains and media

E. coli strain Y1090 R- [Δ*lacU169 proA* Δ*lon araD139 strA supF* (*trpC22::Tn10*)(pMC9); Young and Davis, 1983] was used for λgt11 screening. Yeast strain PEP4D (*a/α his1/+ trp1/+ prc1-126/prc1-126 pep4-3/pep4-3 prb1-1122/prb1-1122 can1/can1*; Jones, 1977) was used to isolate ABF-I.

Bacterial media were made as described by Miller (1972). Yeast cells were grown in YPD (1% yeast extract, 2% peptone, 2% dextrose) as described by Sherman *et al.* (1979).

Reagents and Enzymes

p-Nitro Blue Tetrazolium Chloride (NBT) and 5-Bromo-4-chloro-3-indolyl Phosphate-toluidine Salt (BCIP) were purchased from BioRad Laboratories. Goat anti-rabbit lgG (H+L) alkaline phosphatase conjugate was from BioRad Laboratories. All other enzymes were purchased from United States Biochemical.

Generation of Polyclonal and Monoclonal Antibodies Against ABF-I- ABF-I was purified as previously described (Sweder *et al.*, in press). Polyclonal anti-sera was generated by a 50μg injection into a 4lb female New Zealand rabbit. The rabbit was bled 26 days after the first injection. This was followed by a 25μg boost 4 weeks later. A final 25μg boost was administered 10 days later. 40ml of blood were collected nine days after the final boost, followed by two more bleedings 12 and 15 days after the final boost. Each sample of blood was incubated for 30min in a Corex tube at room temperature and then transferred to 4°C for 30 min. The serum

was decanted and the clot was spun at 15,000rpm in a Sorvall SS34 rotor for 10min to get any additional serum. The yield of serum from 40ml of blood was between 24.4-26.4ml. Serum was stored in aliquots at -20°C.

Mouse monoclonal antibodies against ABF-I were raised in RBF/Dn mice by the following procedure. Two mice were injected three times with 10µg of ABF-I at two week intervals. A final boost of 5µg of ABF-I was given two weeks afterwards. Three days after the final boost one mouse was sacrificed and the spleen cells fused with HL-1 myeloma cells. Hybridoma culture supernatants were screened for the presence of ABF-I antibody using the gel retardation assay described in the preceding chapter. Typically, 33ng of ABF-I were incubated with 1.25ng of ^{32}P-labeled DNA containing the ABF-I recognition sequence and the *ARS* core consensus in the presence of 20-25µg of hybridoma supernatant. Positives clones were subcloned by limiting dilution.

Preparation of Antibody Probe

Anti-*E. coli* antibodies were removed from rabbit antisera by pseudoscreening. A fresh overnight culture of Y1090 cells was grown in L broth containing 0.2% maltose and 100µg/ml ampicillin. 0.2ml of cells was infected with 1x10^5 λgt11 vector and plated onto 150mm L plates using 7.5ml of top agar. The plates were incubated at 42°C for 3 hr and then overlaid with a nitrocellulose filter that had been soaked 10mM IPTG and dried. After incubating at 37°C for 3.5 hr, the filters were removed, washed in TBS (50mM Tris, pH7.5, 100mM NaCl) for 5min, and then blocked with 20% Fetal Calf Serum (FCS) in TBS for 1 hr at room temperature. Each filter was treated with 10ml of anti-ABF-I serum

diluted 1:100 in TBS with 20% FCS and incubated 1 hr at room temperature. The antiserum was collected and incubated with a second set of filters that had been overlaid on the λgt11 L plates after the first set of filters had been removed. The antiserum was collected and this procedure was repeated two more times.

Immunoscreening the λgt11 Genomic DNA Library with Anti-ABF-I Antibody-Y1090 was infected with $1.2x10^6$ phage and plated onto 150mm L plates at a density of $6x10^4$ phage/plate. The plates were incubated at 42°C for 2.25 hr after which time IPTG-treated filters were overlaid on the plates. Plates were incubated with the filters for 4 hr at 37°C. Filters were removed and blocked overnight in TBS containing 5% nonfat dry milk at 4°C. The filters were incubated with 10ml of the diluted primary antibody solution for 3 hr at 4°C. Three sequential 10min washes in TBS containing 1% Tween20 (TTBS) were followed by a 3 hr incubation with goat anti-rabbit antibody conjugated to alkaline phosphatase (BioRad). This secondary antibody was used at a 1:1000 dilution in TBS. The filters were washed three times for 10 min in TTBS and then incubated with the development reagent. Development reagent consisted of 1ml of 15mg/ml NBT in 70% dimethylformamide and 7.5mg/ml BCIP in dimethylformamide mixed into 50ml of 0.1M NaHCO3/1mM MgCl2. Positives appeared as purple spots within 20 min of addition of the development reagents. Color development was stopped by washing twice in distilled H_20.

Plaque Purification of Positive λgt11 clones-Agar plugs containing positive plaques from the primary screening were soaked in 0.7ml of SM

media to elute the phage. Phage were purified by rescreening 2-3 times by the procedure mentioned above. Final verification was determined by screening the positive plaques from the polyclonal screening with murine monoclonal antibodies.

DNA Isolation and Manipulation-DNA was isolated from the positive phage using the LambdaSorb Phage Adsorbent system (Promega Biotech). Phage stocks were prepared from plate lysates as described in Maniatis *et al.* (1980). Phage particles were precipitated from 10ml of phage stock with 0.1ml of LambdaSorb. The precipitate was pelleted and washed twice with 1ml of SM media. Phage DNA was released by resuspending the precipitate in 10mM Tris-HCl, pH7.8, 10mM EDTA and incubating 5 min at 70°C. Following centrifugation, the supernatant was extracted twice with phenol/chloroform and the DNA was ethanol precipitated. The purified phage DNA was resuspended in TE and digested with EcoRI. The 4kbp EcoRI insert was gel purified and cloned into the unique EcoRI site of the Bluescript vector, pBluescript SK (-). Multiple restriction enzyme digestions were performed to generate a restriction map of the inserts.

RESULTS

Specificity of Antibody Binding-The specificity of the rabbit polyclonal antibodies was determined by Western blotting using the relatively crude dsDNA-cellulose protein fractions. As can be seen in Figure 1, the antibodies recognized only two proteins on the Western blot whose molecular weights correspond exactly to the molecular weights for

the two protein bands possessing binding activity on Southwestern blots and on silver-stained gels (see Chapter 2).

Gel Retardation Assays of Hybridoma Supernatant-Supernatants from the hybridoma cultures were assayed for the presence of anti-ABF-I antibodies by gel-retardation (Garner and Revzin, 1981; Fried and Crothers, 1981). Supernatants from hybridomas secreting antibodies against ABF-I should retard the migration of the ABF-I-DNA complex if the antibody does not block the DNA-binding domain of the protein. Should the antibody block the DNA-binding domain or alter the protein conformation of ABF-I, such that it no longer binds, there should be a loss of binding that is readily detectable by gel retardation. All of the hybridomas that tested positive for anti-ABF-I antibodies showed increased retardation; none of the ABF-I antibodies inhibited binding of DNA by ABF-I (Fig. 2). W e isolated 10 different cell lines which secreted ABF-I antibodies by this method.

The specificity of the monoclonal binding to ABF-I was confirmed by Western blots. As can be seen in Fig. 3 nine of the ten supernatents positive for ABF-I antibodies from the gel retardation assays recognized two proteins of apparent molecular weights 130 and 135kDa, the same sizes detected by the rabbit polyclonal sera. This is in exact agreement with the sizes found for the active binding species by Southwestern blotting and silver-stained SDS-PAGE gels (Sweder *et al.*, in press, Chapter 2).

Immunoscreening of λgt11 Yeast Genomic DNA expression Library with Anti-ARS Binding Factor I Antibody-We screened a λgt11 genomic

DNA library using antibodies to *ARS* Binding Factor I (ABF-I) to isolate clones expressing ABF-I epitopes. Screening by the method of Mierendorf *et al.* (1987) with minor modifications we obtained 32 positive clones in the primary screen of $1.2x10^6$ plaques. Plaques corresponding to the positive signals were pulled from the plates and rescreened twice to plaque purify the phage. Ten of the original 32 clones appeared to be positive in all screens. In order to verify that the clones contained the ABF-I gene, all were rescreened using a monoclonal antibody. All 10 phage clones were recognized by the monoclonal antibody, thus confirming the presence of the ABF-I epitope (Fig. 4).

Characterization of DNA Inserts-DNA was prepared from those phage which expressed ABF-I epitopes. The purified DNA was restricted with EcoRI and separated by electrophoresis through a 0.7% agarose gel. Four fragments, ranging in size from 2.3 to 4.4kbp, were gel purified, and cloned into the unique EcoRI site of the Bluescript vector, pBluescript SK (-). Bacterial colonies containing recombinant bluescript plasmids are white while colonies containing only vector are blue. Recombinant DNA was isolated from the white colonies and multiple restriction digests were performed to generate the restriction map shown in Figure 5.

Four of the ten phage contained large inserts, two of which were of sufficient size to encode for the ABF-I gene, based on the molecular weight of the protein. Suprisingly, there is overlap between just two of the DNA fragments on the basis of restriction fragment lengths despite the fact that a single monoclonal antibody recognizes all four of the original λgt11 clones.

DISCUSSION

The gene encoding the *ARS* binding factor I (ABF-I) has been cloned from a yeast genomic DNA expression library in λgt11. The identity of the ABF-I gene was confirmed by two methods. First, monoclonal antibodies against ABF-I recognized the same λgt11 clones as were detected by the polyclonal antibodies. Also, both the monoclonal and polyclonal antibodies retarded the migration of the ABF-I-DNA complex in gel retardation assays, thereby confirming that the antibodies were recognizing the active species.

Four different EcoRI DNA fragments were isolated from the λgt11 clones expressing the ABF-I epitope. As the ABF-I protein is 135kDa, the ABF-I gene should be on the order of 3.7kb. Only two of the λgt11 clones, 3d and 15a, contain a large enough insert to encode the entire protein. The 2.3kb clone, 18a (Fig. 5), shows a similar restriction pattern to the 5′ end of the 4.4kb fragment from 3d. Suprisingly, restriction digests revealed little or no overlap between the other two fragments and fragments 18a and 3d, despite the fact that all of the clones produce proteins which are recognized by a single monoclonal antibody. This may indicate that there has been a rearrangement within two of the clones that would change the restriction pattern without removing the ABF-I epitope. It is also possible that there is a family of antigenically related *ARS* binding factors.

The ABF-I protein has not been directly implicated in yeast DNA replication; however, the role of domain B in plasmid replication suggests a major role for ABF-I. The isolation of the ABF-I gene will allow us to dissect the function of the ABF-I protein *in vitro* and *in vivo*. Deletions can

be made in the cloned gene and used to replace the chromosomal copy of ABF-I, yielding cells which make no functional ABF-I protein (Rothstein, 1983). If this disruption is lethal, indicating that ABF-I performs an essential function, germinating spores should arrest with a terminal phenotype which will indicate at what point in the cell cycle the gene product acts. If ABF-I is not an essential gene, the phenotype of the cells in the complete absence of ABF-I should give us an indication of its function.

If the ABF-I protein is involved in initiation of replication, it is quite likely that it acts in conjunction with other replication proteins. Binding of ABF-I to domain B may be required from the subsequent binding of other proteins, as is the case for the *E. coli* dnaA protein. As element B can be deleted from plasmids without completely abolishing replication, ABF-I may be dispensible. However, these plasmids replicate inefficiently, thus binding of ABF-I may change the DNA conformation at *ARS1* to allow access to other replication proteins. The purified protein does not appear to have either ATPase or helicase activities (Sweder and Campbell, unpublished results); however, sequences adjacent to the ABF-I recognition site have not been tested for unwinding upon binding of the protein. P1 nuclease sensitvity assays (see Umek and Kowalski, 1987, 1988) could be used to address this point.

The role of transcriptional activators in initiation of replication may be through such local unwinding of DNA. Nuclear factor I and nuclear factor III of adenovirus have been shown to greatly enhance binding of the initiation complex (reviewed in DePamphilis, 1988). The SV40 replication origin is flanked by the early enhancer and promoter

sequences. Deletion of the 72bp enhancers has no effect on DNA replication as long as three or more of the six G boxes in the early promoter are present. *In vivo* replication is reduced 2- to 3-fold by deletion of all six G boxes. Removal of all the G boxes and enhancers reduces replication 10- to 100-fold, depending on the other sequences present. A test of this role for ABF-I would be straightforward if the genomic deletion is viable.

ABF-I may instead impose temporal control on the firing of its specific origins. The ability of plasmids lacking the ABF-I binding site to replicate *in vitro* doesn't indicate that such regulation is unaffected. The absence of this protein in replicating cells may allow replicons to be used more than once per S phase, or may change the timing of replication of a given replicon with respect to others. This is difficult to test without a fully competent yeast *in vitro* replication system. This difficulty is compounded by the presence of many origins without ABF-I binding sites. If there is indeed a family of ABF-I-related proteins, as the immunological data might suggest, then cloned genes for three of these proteins may already be in hand.

Details concerning the role of ABF-I in DNA replication are critical to our understanding of the mechanism of replication. Also important is an understanding of the regulation of replication proteins within the yeast cell cycle. For example, the genes encoding DNA polymerase I, DNA ligase, thymidylate synthetase, and thymidylate kinase are coordinately regulated, with their transcript levels peaking just before S phase (Johnston *et al.*, 1987; Budd *et al.*, 1988). Cell cycle regulation of ABF-I at the protein level can be easily tested using gel retardation assays to

determine the presence of ABF-I in extracts made from synchronous cultures of wild type and gene-disrupted strains at intervals throughout the cell cycle. Confirmation can be easily done using the antibodies we have generated in Western blots.

The isolation of ABF-I and several other replication proteins presents an opportunity to create an *in vitro* replication system from purified proteins, similar to those systems already in existence for *E. coli*, SV40, and bacteriophages T4, T7, and λ.

REFERENCES

Baker, T. A., and Kornberg, A. (1988) *J. Biol. Chem.*, in press

Bouton, A. H., and Smith, M. M. (1986) *Mol. Cell. Biol.* **6**, 2354-2363

Brand, A. H., Micklem, G., and Nasmyth, K. (1987) *Cell* **51**, 709-719

Brewer, B. J., and Fangman, W. L. (1987) *Cell* **51**, 463-471

Buchman, A. R., Kimmerly, W. J., Rine, J., and Kornberg, R. D. (1988) *Mol. Cell. Biol.* **8**, 210-225

Celniker, S. E., and Campbell, J. L. (1982) *Cell* **31**, 201-213

Celniker, S. E., Sweder, K. S., Srienc, F., Bailey, J. E., and Campbell, J. L. (1984) *Mol. Cell. Biol.* **4**, 2455-2466

Dean, F. B., Borowiec, J. A., Ishimi, Y., Deb, S., Tegtmeyer, P., Hurwitz, J. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 8267-8271

DePamphilis, M. L. (1988) *Cell* **52**, 635-638

Diffley, J. F. X., and Stillman, B. W. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2120-2124

Fried, M., and Crothers, D. M. (1981) *Nucl. Acids Res.* **9**, 6505-6525

Garner, M. M., and Revzin, A. (1981) *Nucl. Acids Res.* **9**, 3047-3060

Huberman, J. A., Spotila, L. D., Nawotka, K. A., El-Assouli, S. M., and Davis, L. R. (1987) *Cell* **51**, 473-481

Jones, E. W. (1977) *Genetics* **85**, 23-33

Kearsey, S. (1984) Cell **37**, 299-307

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY

Mierendorf, R. C., Percy, C., and Young, R. A. (1987) *Meth. Enzymol.* **152**, 458-469

Miller, A. M., and Nasmyth, K. A. (1984) *Nature* **312**, 247-251

Miller, J. H. (1972) *Experiments in Molecular Cloning*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY

Palzkill, T. G., Oliver, S. G. and Newlon, C. S. (1986) *Nucl. Acids Res.* **14**, 6247-6264

Rothstein, R. J. (1983) *Meth. Enzymol.* **101**, 202-211

Sherman, F., Fink, G. R., and Hicks, J. (1979) *Methods in Yeast Genetics*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY

Snyder, M., Sapolsky, R. J., and Davis, R. W. (1988) *Mol. Cell. Biol.* **8**, 2184-2194

Sweder, K. S., Rhode, P. R., and Campbell, J. L. (1988) *J. Biol. Chem.*, in press

Umek, R. M., and Kowalski, D. (1987) *Nucl. Acids Res.* **15**, 4467-4480

Umek, R. M., and Kowalski, D. (1988) *Cell* **52**, 559-567

Wold, M. S., Li, J. J., and Kelly, T. J. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 3643-3647

Young, R. A., and Davis, R. W. (1983) *Science* **222**, 778-782

FIGURE LEGENDS

Fig. 1 **Immunoblot analysis using rabbit polyclonal antibodies against ABF-I.** Native DNA-cellulose fractions were electrophoresced in a 7.5% polyacrylamide gel and transferred to nitrocellulose. Lane a was reacted with pre-sera. Lanes b and c were reacted with ABF-I antisera from the first and second bleeds, respectively. Lane d was reacted with murine polyclonal ABF-I antisera.

Fig. 2 **Gel retardation assay of hybridoma supernatants.** Hybridoma supernatants were assayed as described in Sweder *et al.* (in press) using the EcoRI-HindIII fragment from YRp09ΔH103. Each binding reaction contained 32.5μl of supernatant in a 45μl reaction volume with 2ng of labeled probe. All reactions contained 1μg of salmon sperm DNA. Free DNA migrates the fastest. ABF-I-DNA complex migrates very slowly at the top of the gel. The antibody-ABF-I-DNA complex migrates the slowest of all the species in the gel.

Fig. 3 **Immunoblot using hybridoma supernatants.** Supernatants from the ten cell lines selected by gel retardation assay were reacted with native DNA-cellulose fractions transferred to nitrocellulose. The leftmost lane was reacted with rabbit polyclonal antisera against ABF-I.

Fig. 4 **Plaque purification of λgt11 clones that react with ABF-I antibody.** Filters from rescreening of two potential positives using ABF-I antibody. The filter on the right was rescreened using murine monoclonal ABF-I antibody.

Fig. 5 **Restriction map of EcoRI inserts from λgt11 clones.** Four restriction fragments were isolated which encoded proteins recognized by the ABF-I antibody. The EcoRI inserts from these clones were subcloned into the unique EcoRI site of pBluescript SK (-).

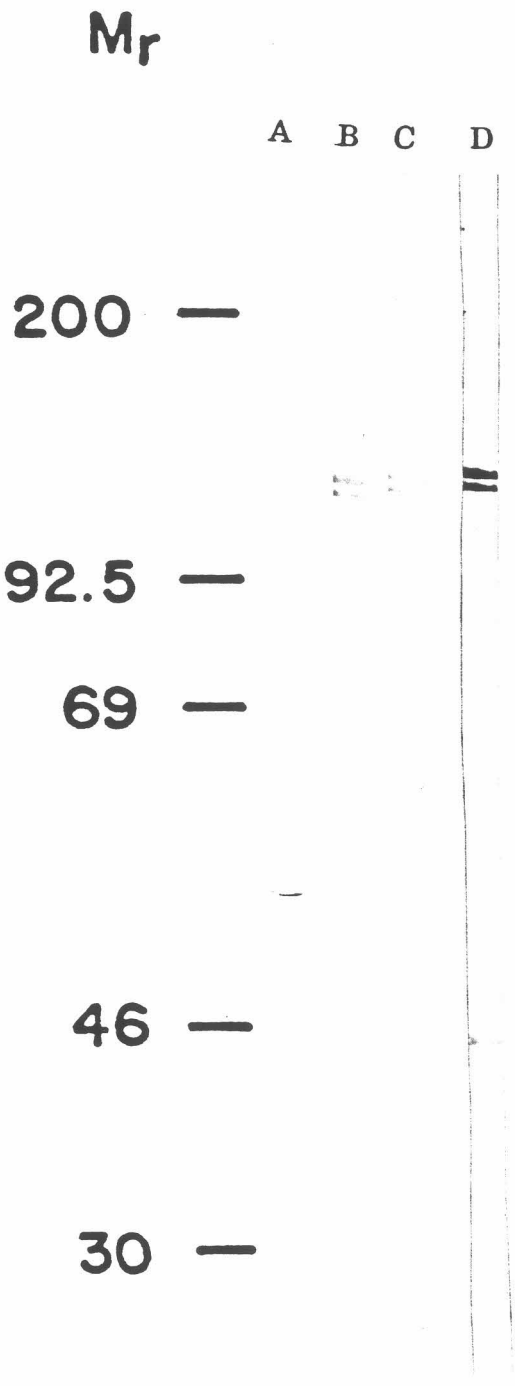


Fig. 1.

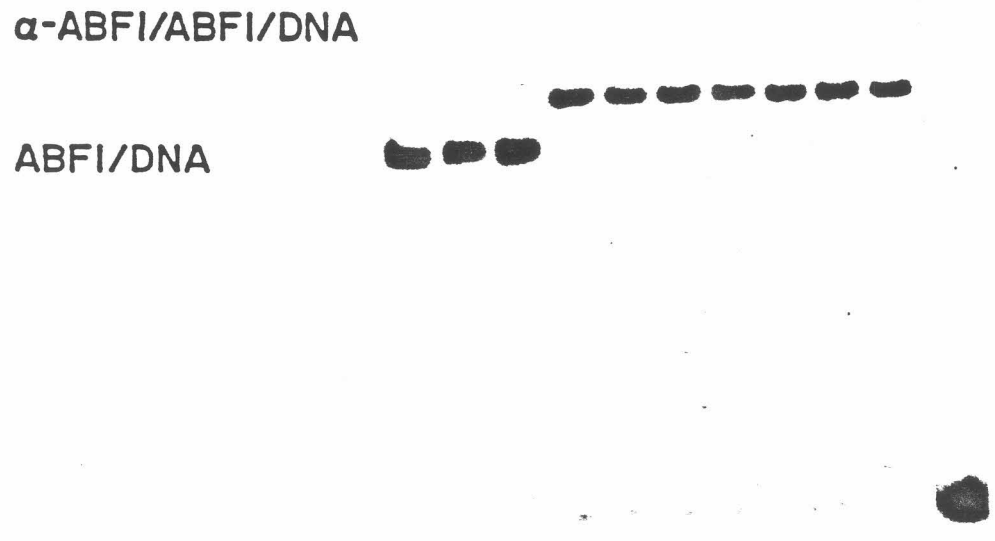


Fig. 2.

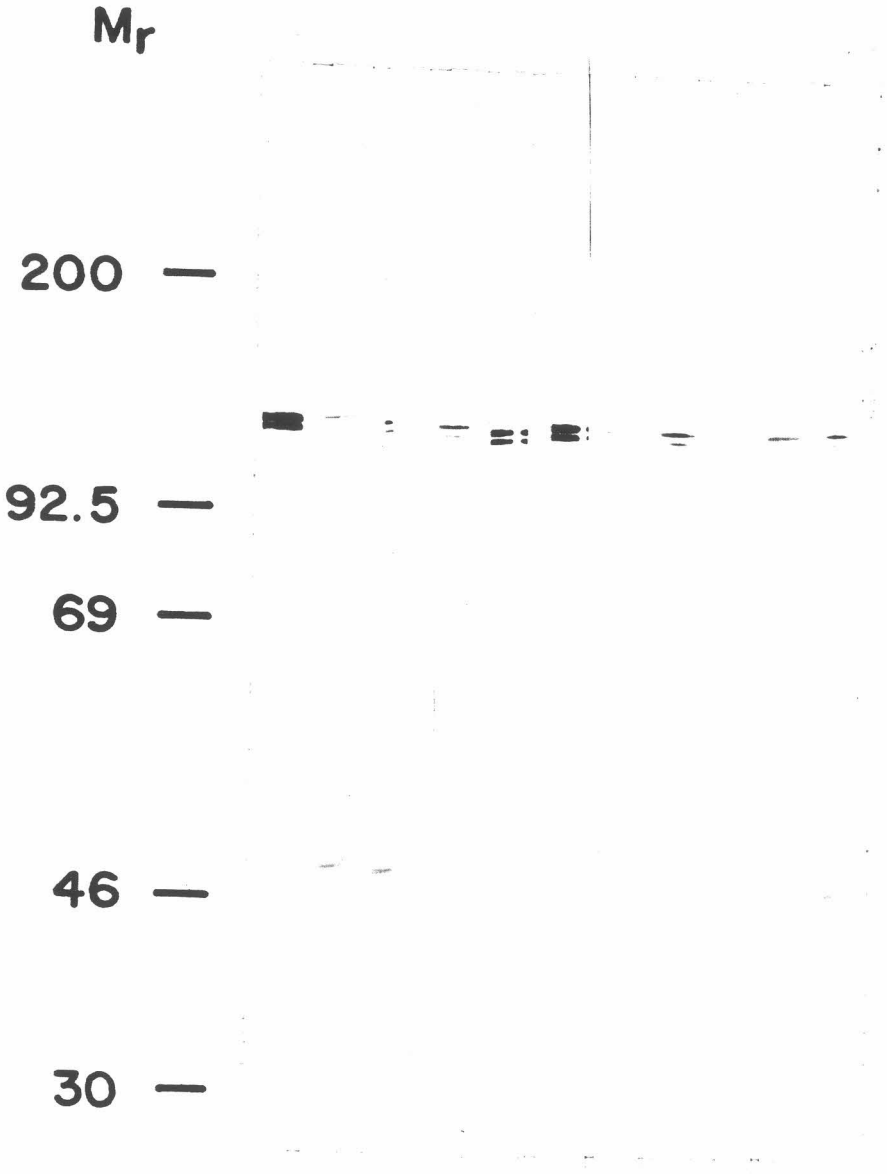


Fig. 3.

Fig. 4.

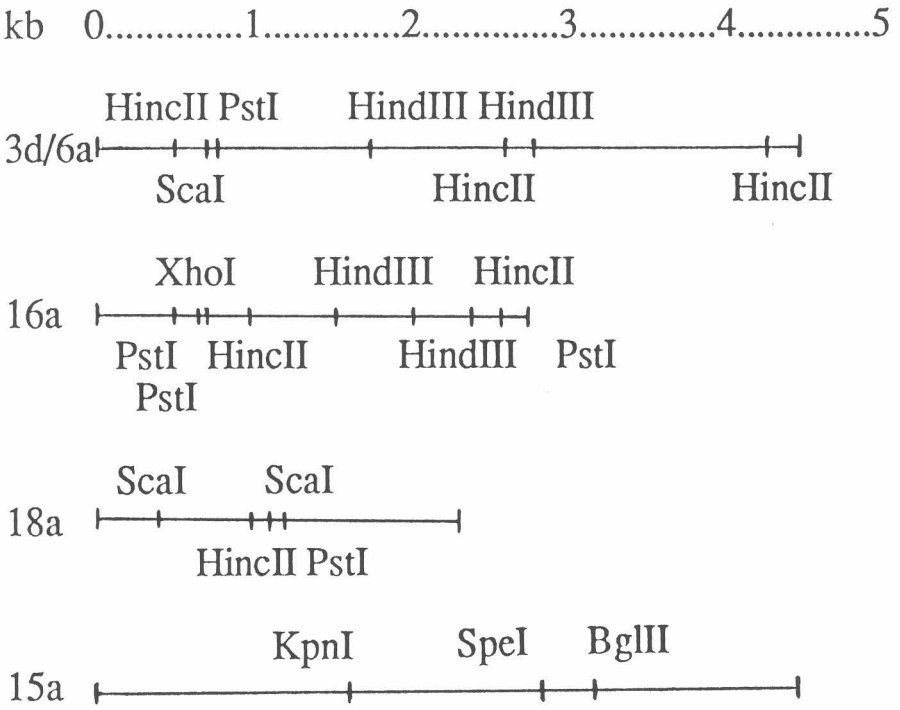

kb 0.....1.....2.....3.....4.....5
HincII PstI
HindIII HindIII
3d/6a
ScaI
HincII
HincII
XhoI
HindIII
HincII
16a
PstI
HincII
HindIII
PstI
PstI
ScaI
ScaI
18a
HincII PstI
KpnI
SpeI
BglII
15a

Fig. 5.