Volume 14 Number 15 1986 Nucleic Acids Research

DNA sequence analysis of *ARS* elements from chromosome III of *Saccharomyces cerevisiae*: identification of a new conserved sequence

Timothy G.Palzkill[1,2], Stephen G.Oliver[3] and Carol S.Newlon[1]

[1]Department of Microbiology, UMDNJ-New Jersey Medical School, 100 Bergen St., Newark, NJ 07103, [2]Genetics PhD Program, University of Iowa, Iowa City, IA 52242, USA and [3]Department of Biochemistry and Applied Molecular Biology, University of Manchester Institute of Science and Technology, PO Box 88, Manchester M60 1QD, UK

Received 3 June 1986; Revised and Accepted 18 July 1986

ABSTRACT

Four fragments of Saccharomyces cerevisiae chromosome III DNA which carry ARS elements have been sequenced. Each fragment contains multiple copies of sequences that have at least 10 out of 11 bases of homology to a previously reported 11bp core consensus sequence (10, 11). A survey of these new ARS sequences and previously reported sequences revealed the presence of an additional 11bp conserved element located on the 3' side of the T-rich strand of the core consensus. Subcloning analysis as well as deletion and transposon insertion mutagenesis of ARS fragments support a role for 3' conserved sequence in promoting ARS activity.

INTRODUCTION

By analogy to prokaryotic systems and eukaryotic viruses, eukaryotic DNA replication is expected to be initiated at specific sequences along the chromosomes termed origins of replication. Candidates for such sequences have been identified in the yeast Saccharomyces cerevisiae and are termed autonomously replicating sequences or ARS's. Although there is as yet no direct evidence for ARS's acting as origins of replication, considerable circumstantial evidence has accumulated (reviewed in 1). Briefly, ARS's, when present on a plasmid, increase the transformation efficiency of the plasmid 10^3-10^4-fold. These plasmids are present in the nucleus of transformed cells in the form of extrachromosomal covalently closed circular DNA (2, 3, 4) and replicate once each cell cycle during S phase under the same genetic control as chromosomal DNA (5, 6). The frequency of ARS's in the genome, one per 32-40 kb, (7, 8) is in agreement with estimates of the spacing of DNA replication origins in vivo obtained in electron microscopic studies (9).

Several ARS elements have been sequenced and all have been found to contain the sequence $\overset{A}{T}TTTAT\overset{A}{G}TTT\overset{A}{T}$ (10, 11). This 11bp core consensus sequence is necessary but not sufficient for ARS function (12, 13, 14). Deletion experiments have demonstrated that sequences 3' to the ARS core consensus are needed to

enable ARS1 (12), the HO ARS (13), and the histone H4 ARS (14) to promote high frequency transformation of the integrating vector, YIp5. A 19 bp fragment containing the ARS1 core consensus can promote high frequency transformation of a centromere-containing vector despite its inability to act in YIp5. However, even in the CEN vector, fragments larger than 19 bp yield transformants that are more stable (15). On the basis of their analysis of deletion mutants, the Campbell group has proposed that ARS1 consists of three domains. Domain A is the 11bp core consensus; domain B is 3' to the T-rich strand of the core consensus; domain C maps 5' to the core consensus. When inserted into YIp5, deletions into domain B result in plasmids which are less stable than plasmids containing the intact ARS. The effects of deletions into domain C can be seen only in the absence of domain B (12), suggesting that domain B is more important for ARS function. When assayed in a centromere-containing vector, the effects of deletions into either domain B or domain C are evidenced by a reduction in plasmid stability with deletions in domain B having a stronger effect (15). The nature of the requirement for 3' sequences, corresponding to domain B, is unknown since no DNA sequence homologies among these sequences have previously been identified.

We report here the DNA sequences of four ARS elements derived from a 200 kb circular derivative of chromosome III which was formed by homologous recombination between the HML locus on the left arm and the MAT locus on the right arm of the chromosome (16). The ring chromosome has been cloned and has been found to contain 12 ARS's (17 and Ms in prep). Our analysis of four of these ARS's (C2G1, A6C, H9G and J11D1), has revealed that all contain several close matches to the previously reported 11 bp core consensus sequence (10, 11). In addition, our survey of these new ARS sequences and of previously reported ARS sequences permitted us to identify DNA sequence homologies located on the 3' side of the T-rich strand core consensus that appear to represent an additional conserved element of ARS sequence.

MATERIALS AND METHODS

Strains, media

E. coli strain JA226 (18) was used as the host for large scale isolation of plasmid DNA for sequencing. JA226 transformants were grown in Luria Broth (19) supplemented with 50 µg/ml ampicillin. E. coli strains HB101 carrying F'8 lac pro and XK1502 (F^- Δ lac Nal^r), used in transposon γδ mutagenesis were obtained from Dr. John Woolford. S. cerevisiae strain YNN27 (MATα trp1-289 ura3-52 gal 2) was used as the recipient for ARS assays (20). Transformants were selected on -Ura plates which contained per liter 6.7 gm yeast nitrogen base without amino acids, 10

gm casamino acids, 10 mg adenine, 10 mg tryphophan, 20 gm glucose and 20 gm agar. Growth curves on transformants were measured at 30°C in liquid medium of the same composition from which agar was omitted.

Plasmid structure

Chromosome III fragments chosen for sequencing were 500-900 bp fragments demonstrated to have ARS activity when subcloned in the integrative plasmid YIp5 (Newlon et al., in preparation). C2G1 is a 521 bp EcoRI-ClaI fragment approximately 5 kb distal to CEN3 on the left arm of the chromosome (plasmid pC2G34c). A6C is a 554 bp Nru-ClaI fragment which maps approximately 25 kb distal to HIS4 on the left arm (plasmid pA6CB6). J11D1 is an 860 bp SalI-EcoRI fragment which maps approximately 5 kb centromere proximal to PGK1 on the right arm (plasmid pJ11D5a1) and H9G is a 852 bp EcoRV fragment which maps approximately 9 kb centromere proximal to CRY1 on the right arm (plasmid pH9Ga2RV). The H9G fragment was cloned into YIp5 using BamHI linkers. The other three fragments were cloned between sites corresponding to their ends in YIp5. As a control for possible plasmid context effects in the C2G1 subcloning studies reported here, the 521 bp EcoRI-ClaI fragment containing the C2G1 ARS was inserted into the BamHI site of YIp5 using BamHI linkers (plasmid pC2G1B). Subclones of this fragment were obtained by adding BamHI linkers to restriction fragments whose sticky ends were filled in by the Klenow fragment of DNA polymerase I and inserting the fragments in the BamHI site of YIp5. The orientation of sequences in the figures corresponds to their orientation in chromosome III with LEU2 and HIS4 on the left arm as is the standard convention.

Transposon insertion mutagenesis

The γδ transposon, a relative of Tn3, is carried on F' plasmids. Transposition of γδ into a pBR322-related plasmid produces a cointegrate intermediate which can be transferred to an F' strain by virtue of the F' transfer functions. The intermediate then resolves in the transconjugant, yielding pBR322 plasmid derivatives carrying the transposon (23). To select transposon insertions in the C2G1 fragment, strain HB101 carrying the F'8 plasmid was transformed with pC2G34c. Transformants were selected and cells from single colonies were mated to strain XK1502. Transconjugants carrying pC2G34c with an integrated γδ transposon were selected on LB plates containing 50 μg/ml ampicillin and 150 μg/ml nalidixic acid. Several colonies were picked from each mating and plasmid minipreps were made. The position and orientation of the transposon was mapped using BamHI and EcoRI which cut both in the transposon (24) and in pC2G34c. Fine structure mapping used the SacI site at the terminus of γδ and a variety of sites

within the C2G1 ARS fragment. Mapping of insertion sites is accurate to within ± 10 bp.

Plasmid isolation

Plasmid DNA was prepared from E. coli according to the alkaline lysis procedure as described in (21). The DNA was purified in cesium chloride gradient by either equilibrium centrifugation or using a step gradient method (22). Plasmid minipreps were made from 1.5 ml cultures using the alkaline lysis protocol (21).

In vitro labelling.

Restriction endonuclease digested DNA's were fractionated on 8% polyacrylamide gels and electroeluted in dialysis tubing (21). Restriction fragments were 5' end labelled using T4 polynucleotide kinase after treatment with bacterial alkaline phosphatase (25). 3' end labelling of restriction fragments was performed using the Klenow fragment of DNA polymerase I (25).

DNA sequencing.

The DNA sequencing of the A6C, J11D1 and H9G fragments was performed using the Maxam and Gilbert chemical cleavage method (25). The C2G1 fragment was sequenced using a combination of the chemical cleavage method and the Sanger chain termination method (26). Sanger sequencing was performed on M13 subclones of the C2G1 ARS fragment, made as described previously (27).

Computer analysis of DNA sequences.

Restriction enzyme sites and open reading frames in the sequenced ARS's were detected using the PCS DNA Sequence Analysis package (27). DNA homology searches were performed on a Hewlitt Packard A900 computer. The algorithms were essentially those described by Staden (29,30), implemented by G.R. Cleaves in Fortran 77. These programs are available on request. DNA sequences other than the chromosome III ARS's reported here and the $tRNA^{Glu}$ - containing fragment (31) were obtained from the GenBank data base supplied by Bolt, Beranek and Newman, Inc.

RESULTS

As an approach to determining DNA sequence requirements for ARS function, we have begun a study of ARS's on yeast chromosome III. The ARS-containing fragments whose sequences are presented here are 500 to 900 bp in length. Plasmids containing these small fragments transform S. cerevisiae strains as efficiently as plasmids carrying ARS1 or the longer chromosome III fragments from which they were derived (200-400 transformants/ µg DNA using the lithium acetate transformation procedure). In addition, these plasmids are extrachromosomal and mitotically unstable as expected for ARS-containing plasmids (Newlon et al., in preparation).

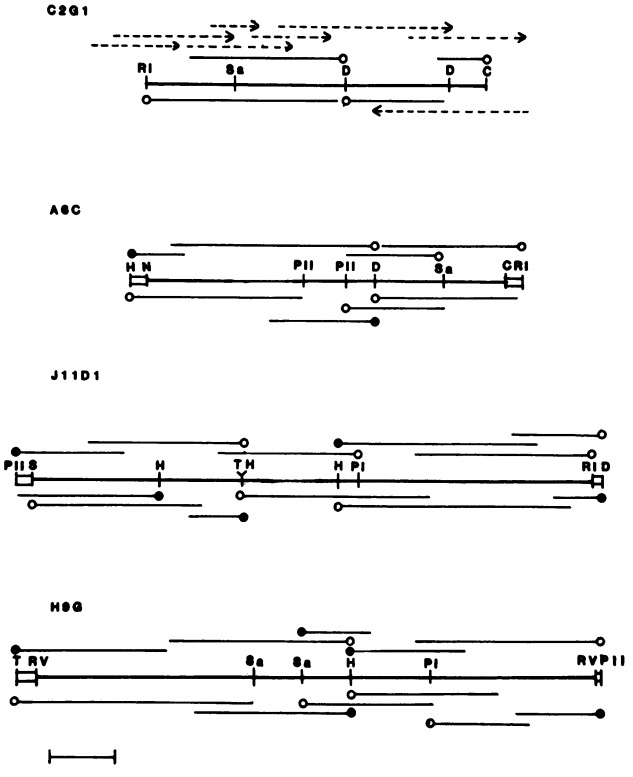


Figure 1. Sequencing strategies for chromosome III ARS-containing fragments. Abbreviated restriction maps of the fragments are shown in heavy lines with vector sequences indicated by boxes. Dashed lines indicated sequences obtained by the chain termination method. Solid lines indicated sequences obtained by the chemical cleavage method. Open circles represent 3' end label and solid circles, 5' end label. C, ClaI; D, DdeI; H, HinfI; N, NruI; PI, HpaI; PII, HpaII; RI, EcoRI; RV, EcoRV; S, SalI; Sa, Sau3A1; T, Tag I. (The bar represents 100 bp.)

The sequencing strategies developed for the chromosome III ARS-containing fragments are diagrammed in Figure 1. The DNA sequences obtained are presented in Figures 2-5.

The ARS core consensus

As an initial step in attempting to delineate sequences required for ARS function, the sequences were scanned for near matches to the known ARS consensus, $\begin{matrix}A\\T\end{matrix}TTTAT\begin{matrix}A\\G\end{matrix}TTT\begin{matrix}A\\T\end{matrix}$. The results of this analysis are diagrammed in Figures 2-5. Only one of the four chromosome III ARS-containing fragments, C2G1, contains a perfect match to the 11 bp consensus (Fig. 2). However, all four of the chromosome III ARS's have a high density of near matches. The A6C ARS contains

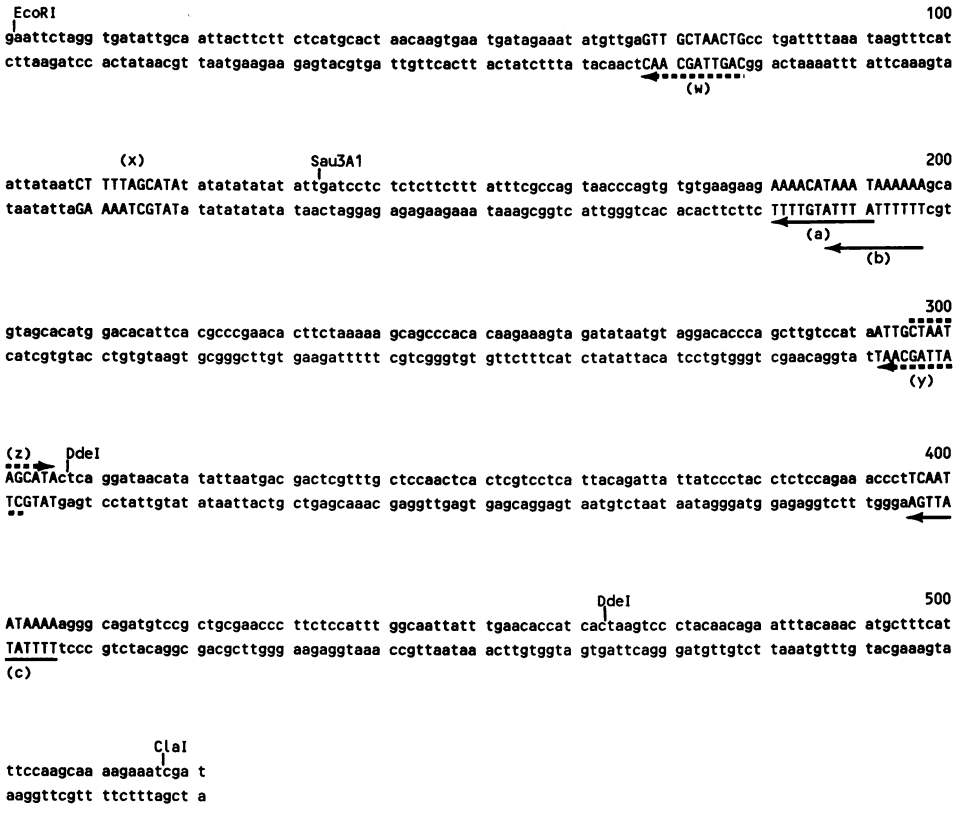


Figure 2. C2G1 sequence. Terminal restriction sites and restriction sites used for subcloning are shown. Matches to the ARS core consensus (11 out of 11 and 10 out of 11) are shown in capital letters with the T-rich strand underlined or overlined with a solid arrow. Matches are lettered sequentially beginning with (a). Matches to the 3' conserved sequence are shown in capital letters and are underlined or overlined with a dotted arrow. They are lettered with letters at the end of the alphabet (w-z).

four 10 out of 11 matches and seven 9 out of 11 matches (Fig. 3). The J11D1 ARS contains four 10 out of 11 and twelve 9 out of 11 matches (Fig. 4). The H9G ARS contains three 10 out of 11 and five 9 out of 11 matches (Fig. 5). In addition to the single 11 out of 11 match, the C2G1 ARS contains two 10 out of 11 and six 9 out of 11 matches. In the case of the J11D1 fragment, further subcloning has localized the ARS activity to a HpaI-EcoRI fragment of 340 bp; similarly, the A6C ARS activity is localized in a 460 bp NruI-Sau3A1 fragment. In each case the fragment with ARS activity contains all of the 10 out of 11 matches and the majority of the 9 out of 11 matches.

Table 1

ARS Core Consensus	Sequence											Reference
	1	2	3	4	5	6	7	8	9	10	11	
HMR (right)	A	T	T	a	A	T	A	T	T	T	A	11
HO	T	T	T	a	A	T	A	T	T	T	T	13
rDNA	g	T	T	T	A	T	G	T	T	T	T	29
rDNA	T	T	c	T	A	T	A	T	T	T	T	29
rDNA	T	T	c	T	A	T	G	T	T	T	T	29
tRNA ^{Glu3}	T	T	T	T	A	T	A	T	g	T	T	30
C2G1a	A	T	T	T	A	T	G	T	T	T	T	This paper
C2G1b	T	T	T	T	t	T	A	T	T	T	A	" "
C2G1c	T	T	T	T	A	T	A	T	T	g	A	" "
A6Ca	T	T	T	T	t	T	A	T	T	T	A	" "
A6Cb	T	T	a	T	A	T	G	T	T	T	T	" "
A6Cc	T	T	T	T	A	T	A	T	g	T	T	" "
J11D1a	A	T	c	T	A	T	A	T	T	T	T	" "
J11D1b	T	T	a	T	A	T	G	T	T	T	A	" "
J11D1c	T	T	T	T	A	T	A	T	g	T	T	" "
J11D1d	T	T	T	T	t	T	A	T	T	T	T	" "
H9Ga	A	T	T	T	A	T	G	T	T	a	T	" "
H9Gb	A	T	T	T	A	c	A	T	T	T	A	" "
H9Gc	T	T	T	c	A	T	G	T	T	T	A	" "

All sequences within a fragment shown to contain an ARS which are homologous to the ARS core consensus (10, 11) in at least 10 out of 11 positions are shown. Lower case letters in the sequence indicate nucleotides which differ from the core consensus. The chromosome III consensus sequences are identified by letters (a-d) which correspond to the sequences highlighted in Figs. 2-5.

Other ARS's have been identified which do not contain an exact match to the core consensus sequence. The HO ARS (13), the HMR (right) ARS (11), the rDNA spacer region (32) and an ARS associated with the tRNA^{Glu} gene (31) lack an exact match to the consensus sequence. Comparison of the 10 out of 11 matches from the above ARS's with the 10 out of 11 matches from the chromosome III ARS's reveals that many of the 10 out of 11 matches vary in the same positions (Table 1). At least one of the 10 out of 11 matches from each of the chromosome III ARS's differs in one of the same positions as the HO, HMR (right), and the rDNA ARS's (positions 3 and 4). This may indicate that positions 3 and 4 (Table 1) can vary and still retain ARS activity. However, it is not known which 10 out of 11 match(es)

promote ARS activity in the chromosome III ARS's or if only a single consensus is required (see Discussion).

A new 3' conserved sequence

In an effort to identify conserved sequences other than the core consensus sequence, comparisons were made of the four chromosome III ARS fragments along with published sequences for fragments containing ARS1 (33), ARS2 (34), and the 2 μm plasmid ARS (35, 36). Previous searches for homologous sequences among ARS's have not detected strongly conserved sequences besides the core consensus sequence. Therefore, it was reasoned that if homologies exist they must be weak. To detect weakly conserved sequences a computer search was used to detect sequences at least 9 bp in length with at least 75% homology. When comparing two ARS's this yielded, on average, 300 matches. Since it is known from studies of other cis-acting elements, including the E. coli replication origin, oriC (37), prokaryotic promoters (38), eukaryotic RNA polymerase III promoters (39), and eukaryotic TATA boxes (40, 41), that spacing relationships among conserved sequences are functionally important, the matches detected in pairwise comparisons of ARS's were screened for those which met two further criteria. First, their position relative to the known consensus sequence had to be constant within 20 bp. In the case of the chromosome III ARS's, all 10 out of 11 matches to the ARS consensus were tested. Second, the sequence was required to be present in three or more ARS's. The most striking homology revealed by these comparisons is a 11 bp sequence found 3' to the T-rich strand of the ARS core consensus.

The 3' conserved sequence has the consensus 5'CTTTTAGC$\overset{A}{T}\overset{A}{T}\overset{A}{T}$3' and is found 55-88 bp 3' to a core consensus in ARS1, ARS2, and all four chromosome III ARS's (Table 2). A match to the element was not found in the 2 μm ARS. The ARS-containing fragments used in the initial comparison were then searched specifically for additional copies the 3' conserved sequence (Table 2). In ARS2, the only copy of the sequence is the one identified in the initial search and it is located 59 bp 3' to the core consensus, between positions 1289 and 1298 of the published sequence (34). In ARS1, two copies were found, the first located 88 bp 3' to the core consensus between positions 762 and 772 of the published sequence, and the second located between positions 894 and 904, approximately 30 bp 3' to a 10 out of 11 match to the core consensus sequence (33). The first 3' conserved sequence in ARS1 lies at the distal end of domain B, a region defined by deletion experiments to be important for ARS function (12). The orientation of the 3' conserved sequence with respect to the core consensus is opposite in ARS1 and ARS2, suggesting that it may function in either orientation.

Three of the chromosome III ARS's have more than one copy of the 3'

Table 2

3' Conserved Element	Sequence										
C2G1 w	C	a	g	T	T	A	G	C	A	A	c
* C2G1 x	C	T	T	T	T	A	G	C	A	T	A
C2G1 y	C	T	a	T	T	A	G	C	A	A	T
C2G1 z	C	T	a	a	T	A	G	C	A	T	A
* A6C x	C	T	T	a	a	A	G	C	T	A	A
A6C y	a	c	c	T	T	A	G	C	T	T	T
A6C z	g	T	T	T	T	A	G	g	A	T	A
* J11D1	g	T	T	T	T	A	G	C	T	c	T
H9G x	t	T	c	T	T	A	G	C	T	T	c
H9G y	C	T	T	T	T	A	G	C	T	c	T
* H9G z	t	c	T	T	T	A	a	C	A	T	c
* ARS1 y	t	T	c	T	T	A	G	C	A	T	T
ARS1 z	g	a	T	T	T	A	G	C	A	T	T
* ARS2	C	T	T	T	T	A	G	C	T	T	c
Consensus	C	T	T	T	T	A	G	C	A/T	A/T	A/T
Number of matches	7	10	8	12	13	14	13	14	14	12	10

The sources of DNA sequences are indicated in Table 1. Capital letters indicate matches to the consensus and lower case letters indicate nucleotides which differ from the consensus. Asterisks identify the sequences identified in the initial pairwise comparisons of ARS sequences.

conserved region. In C2G1 (Figure 2) the conserved sequence is present 4 times with one copy (C2Gx) located 61 bp 3' to the exact match to the core consensus sequence (C2G1a), a second copy (C2Gw) 102 bp 3' to the same core consensus, and with the other two copies (C2Gy and C2Gz) located as overlapping inverted repeats approximately 90 bp 3' to a 10 out of 11 match to the ARS consensus (C2G1c). In the A6C ARS, (Figure 3) the conserved sequence is present three times, twice in the form of overlapping inverted repeats (A6Cx and A6Cy) which are located approximately 70 bp 3' to a pair of overlapping matches to the core consensus sequence (A6Cb, A6Cc), and a single copy (A6Cz) which is 51 bp 3' to a single core consensus (A6Ca). In J11D1 (Figure 4) the conserved sequence is present only once, approximately 70 bp 3' to a pair of overlapping matches to the core consensus (J11Db, J11Dc). In the H9G ARS (Figure 5), the conserved sequence is present three times with one copy (H9Gx) located 112 bp 3' to core consensus H9Gb the second copy (H9Gy) located 103 bp 3' to core consensus H9Ga, and the third copy (H9Gz) located 55 bp 3' to core consensus H9Gc.

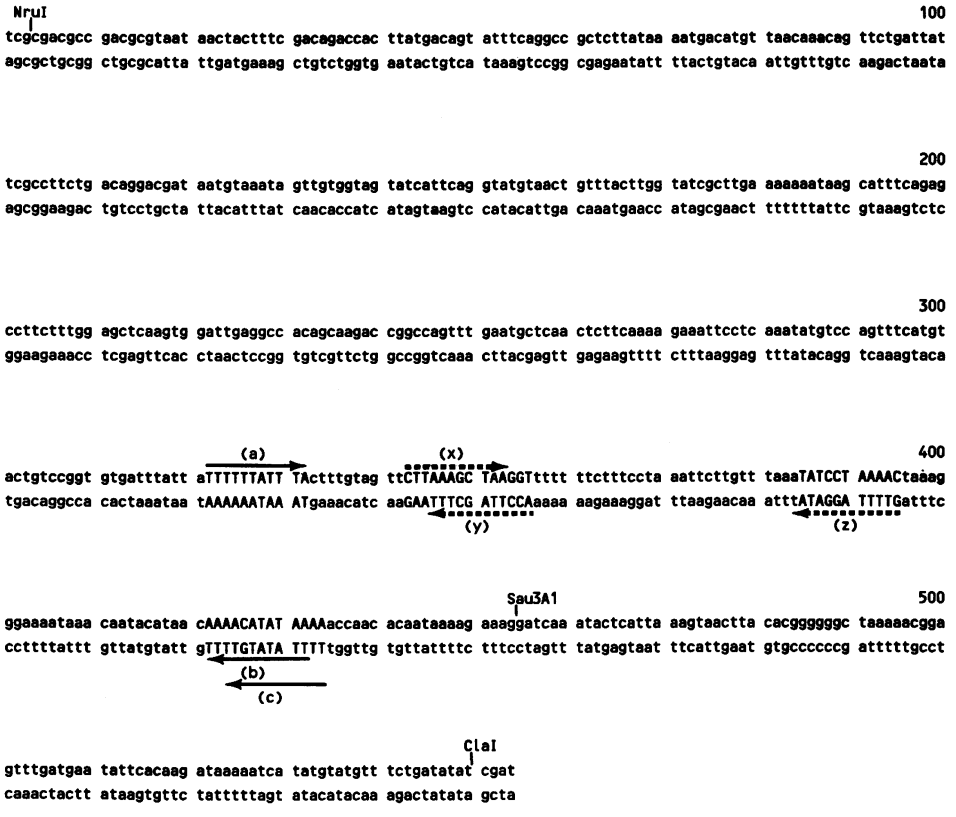


Figure 3. A6C sequence. Restriction sites and symbols as described in Figure 2.

Other published ARS sequences were searched for the 3' conserved sequence. There is no similar sequence in the histone H4 ARS, but there is limited published sequence 3' to the core consensus (14). Close matches to the sequence are found in the HO ARS (13), the $tRNA^{Glu}$ fragment (31) and the HMR (right) fragment (11), but their spacing relative to the core consensus is outside the initial criteria. In the HO ARS, the near match is within 30 bp of the core consensus and in the other two ARS's, the 3' conserved sequence is more than 100 bp away from the core consensus. Thus six of eleven ARS-containing sequences examined have a match to the 3' conserved sequence between 55 and 88 bp 3' to their core consensus sequence(s). Three other ARS's have copies of the sequence but at closer or more distant spacings.

Subclones and transposon insertions in C2G1 ARS with impaired function.

To further assess the significance of the 3' conserved element and to

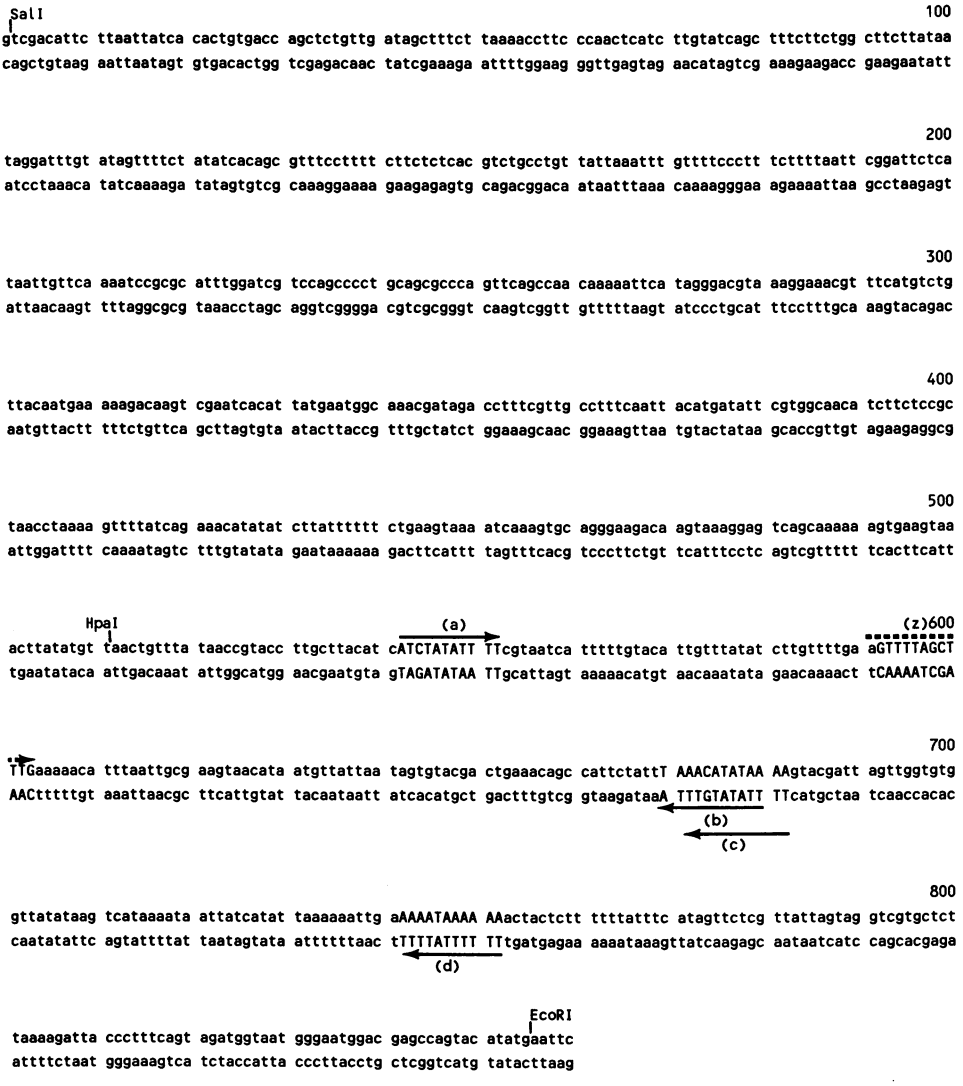


Figure 4. J11D1 sequence. Restriction sites and symbols as in Figure 2.

determine which of the matches to the known consensus contribute to ARS function, the C2G1 ARS was subcloned. The 521 bp EcoRI-ClaI fragment was divided into three fragments: a 137 bp EcoRI-Sau3A1 fragment, a 176 bp Sau3A1-DdeI fragment and a 155 bp DdeI-DdeI fragment, and each fragment as well as the 521 bp EcoRI-ClaI fragment was inserted into the BamHI site of YIp5 in both orientations. When the resulting plasmids were tested for ARS activity using the

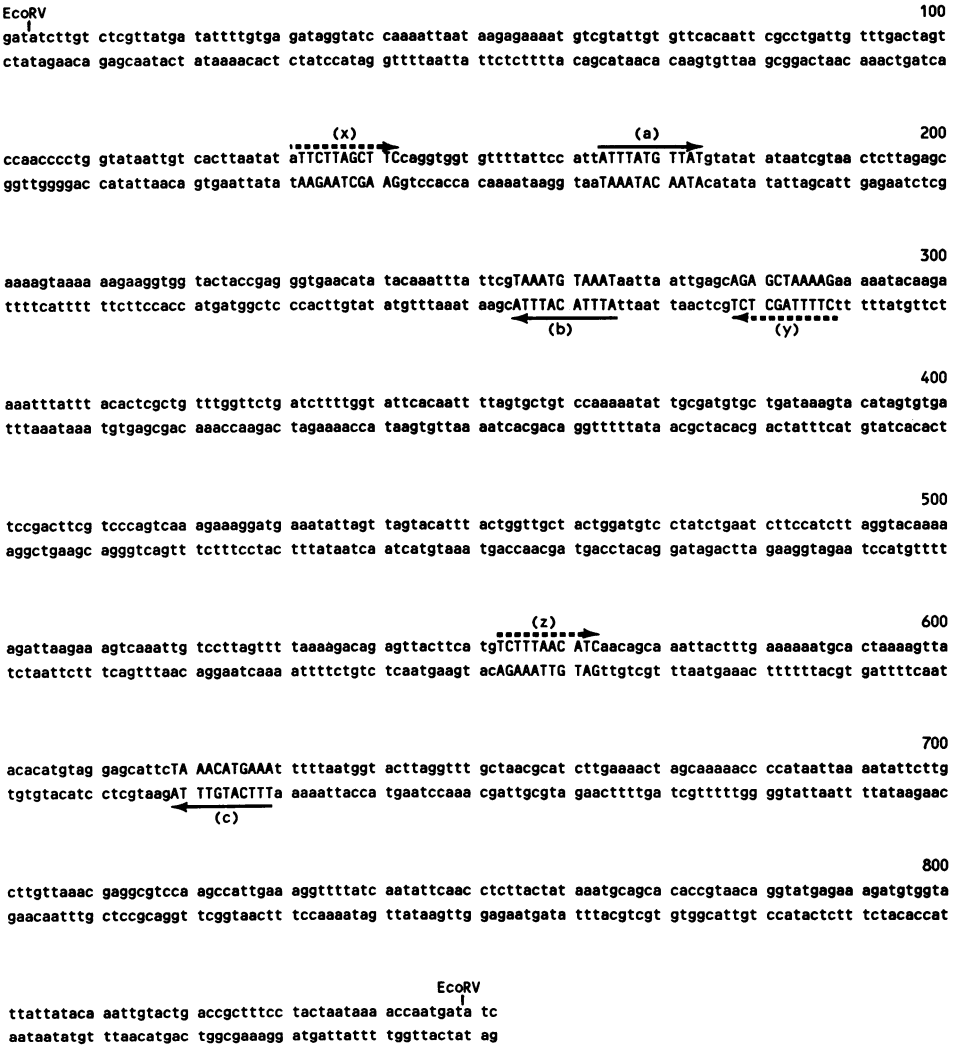
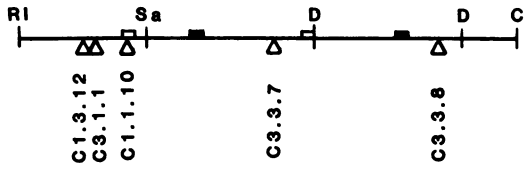


Figure 5. H9G sequence. Restriction sites and symbols as in Figure 2.

high frequency transformation assay, the plasmids containing the 521 bp C2G1 fragment transformed strain YNN27 as efficiently as pC2G34c (200-400 transformants/ µg DNA), while none of the subclones transformed more efficiently than YIp5 (0-1 transformants/ µg DNA, data not shown).

The 170 bp Sau3A1-DdeI fragment contains the exact match to the core consensus (C2Ga) along with 46 bp of DNA flanking the 3' side of the consensus and 119 bp flanking the 5' side of the consensus sequence, while the DdeI-DdeI fragment



Plasmid	Doubling Time (hrs)
pC2G34c	2.25
C1.3.12	3.4
C3.1.1	3.4
C1.1.10	15.0
C3.3.7	2.4
C3.3.8	2.4

Figure 6. Transposon insertions in the C2G1 ARS-containing fragment. The upper part of the figure shows the positions of five γδ insertion in the C2G1 ARS fragment carried in YIp5. The positions of ARS core consensus sequences are shown by solid boxes and of 3' conserved sequences by open boxes. Arrows below the line indicate positions of transposon insertions. The lower part of the figure shows doubling times of transformants grown in selective medium (without uracil) and assayed by measuring the A_{600} of cultures using a Gilford Spectrophotometer.

has 88 bp of 3' sequence flanking a 10 out of 11 match (C2Gc) along with 61 bp of 5' flanking sequences (Figure 2). These results indicate that a core consensus with substantial flanking DNA is not sufficient for ARS activity. These results are also consistent with a role for the 3' conserved sequence in ARS activity since the subcloning separates both of the core consensus sequences from their 3' conserved elements (see Fig. 2). Since the Sau3A1-DdeI fragment contains a copy of the 3' conserved element 5' to the core consensus, these results suggest that the position of the element relative to the core consensus is important.

As an additional approach to identifying regions of functional importance in the C2G1 ARS, transposon insertion mutations were obtained using the γδ element from the F' factor of E. coli (23, 24). Five different insertion mutants were analyzed for transformation efficiency and for growth rate (Fig. 6). The doubling time, under selective conditions, of transformants carrying ARS plasmids provides a qualitative indication of ARS function, with the increased doubling times reflecting increased plasmid loss (10,12,13). Plasmids containing each of the insertion mutations transformed strain YNN27 with the same efficiency as the parent plasmid (200-400 transformants/μg DNA). However, three of the insertion mutants yielded transformants which grew noticably more slowly than transformants containing the parent plasmid. Insertions approximately 75 bp and 240 bp 5' to C2Ga, the exact match to the core consensus sequence, (C3.7.7 and C3.3.8) were

found to have a minimal effect on doubling time of the transformants. Transformants carrying insertions 3' to the C2Ga core consensus have increased doubling times. As would be expected if the increased doubling times reflect increased plasmid loss, the fraction of cells carrying plasmid under selective conditions is reduced in the insertion mutants with longer doubling times. For transformants carrying the parent plasmid grown under selective conditions (-ura medium), 5-6% of cells in the population contain plasmid. This value is similar to that reported for an ARS1 plasmid (12). For transformants carrying plasmids with 3' insertions (Fig. 6), this fraction is less than 1%. These results indicate that sequences 3' to the exact match to the core consensus sequence are important for C2G1 ARS function. It is of interest that the insertion mutation showing the greatest increase in doubling time maps close to or within the 3' conserved sequence. These results also suggest that the 3' conserved sequence and regions around it are important for ARS function.

Additional DNA sequence analysis

The chromosome III ARS sequences were also analyzed for open reading frames. The longest open reading frames found were 60-65 amino acids long and none of the open reading frames longer than 30 amino acids begin with methionine as would be expected for a protein. These results are consistent with analyses of other ARS's. Of the fourteen ARS's whose positions are mapped with respect to known RNA transcripts, only one is in a transcribed region (Newlon, unpublished compilation).

The pairwise comparison of ARS sequences often revealed homologies of significant length which were found only in a particular pair of ARS's. For example, the H9G fragment and the C2G1 fragment have two 16 bp stretches of close homology. Similarly the C2G1 ARS and the HMR (right) ARS share a 16 bp tract of alternating AT. A similar homology between ARS1 and ARS2 was described previously (34). The biological significance of these homologies is currently impossible to assess.

DISCUSSION

We report here the sequences of four chromosome III ARS's, C2G1, A6C, J11D1 and H9G. Each of these ARS's have 3 or 4 copies of a sequence that has at least 10 bases of homology to a previously reported 11bp core consensus sequence (10, 11); only one ARS has an exact match to the 11bp consensus. Two of the ARS's (J11D1 and A6C) have been subcloned and in each case, fragments that retain ARS activity contain the core consensus sequences.

The necessity for at least 1 copy of a core consensus is well established by

deletion analysis of ARS1 (12), the HO ARS (13) and histone H4 ARS (14). The influence of multiple core consensus sequences is as yet unclear. We do not yet know whether more than one consensus sequence is used when more than one is present, nor do we know whether the chromosome III ARS's display enhanced efficiency because of the presence of multiple copies of core consensus. The work of Fangman, et al., suggests that increased copies may enhance the efficiency since four tandem copies of a 64bp fragment of mitochondrial DNA which contains near matches to the core consensus are insufficient for ARS activity while eight tandem copies are sufficient (42).

Arguments for the necessity of DNA sequences in addition to the core consensus sequence for full ARS function come from subcloning experiments. The smallest fragments of ARS1 (12) and ARS2 (34) which have ARS activity in YIp5 are approximately 100bp long. The subcloning experiments reported here also demonstrate that core consensus sequences contained in fragments of 155 bp and 170bp are insufficient to exhibit ARS activity when cloned in either orientation at the BamHI site of YIp5. Further arguments for the necessity of DNA sequences in addition to the core consensus can be made. Searches of yeast DNA sequences known not to have ARS activity revealed the presence of core consensus elements. The region immediately 3' to the copy I histone H3 gene does not have ARS function but does contain an exact match to the core consensus. Similarly, the region between the histone H3 and the histone H4 genes at the copy I locus has two 10 out of 11 matches and the region 3' to the histone H4 gene at the copy II locus has a 10 out of 11 match (43). Finding core consensus sequences without ARS activity in noncoding regions where ARS s are usually found suggests the need for additional sequence elements.

Our search for additional conserved sequences which might provide the additional sequences needed for ARS function revealed an 11bp sequence that is associated with most ARS's and which lies 3' to the core consensus sequence. Assessment of the biological significance of the element will require a detailed genetic analysis. However, a number of features of the element are consistent with its proposed role.

First, the spacing of the element with respect to a core consensus sequence is relatively constant (80 $\pm$ 30 bp). This is similar to the organization of the SV40 origin of replication which has a core region essential for function and flanking DNA regions which increase the efficiency of replication (44). The spacing of the flanking regions relative to the core region is important for the efficiency of origin function (45). By analogy, yeast ARS's contain a core consensus which is absolutely required for function and flanking sequences which increase the efficiency of ARS

function. Since the 3' element we have identified is not found in all ARS's examined, there may be analogous elements which function in other classes of ARS's.

Second, analyses of both the C2G1 ARS and ARS1 support a role for the 3' conserved element in ARS function. Subclones of the C2G1 ARS which separate the 3' conserved sequences from their respective core elements inactivate ARS function and transposon insertions in or near one of the elements reduce the growth rate of transformants under selective conditions. In ARS1, the deletions which define domain B remove the 3' conserved element. These domain B deletions have phenotypes similar to the transposon insertions which are 3' to the core consensus (12). Thus the conserved element we have identified on the basis of DNA sequence analysis may correspond to domain B. Whether the transposon insertions 5' to the core consensus which appear to slightly reduce the growth rate of transformants corrspond to domain C mutations (12,15) remains to be determined.

Third, the 3' conserved element offers an attractive explanation for previous observations of the behavior of the histone H4 ARS in various plasmid contexts (14). When the H4 ARS fragment is inserted between the EcoRI and BamHI sites of YIp5, 66bp of 3' flanking yeast sequence is required for ARS activity, but when inserted in the opposite orientation only 12bp of 3' yeast sequence is required. In this latter orientation, pBR322 DNA sequences provide three near matches to the 3' conserved sequence described here within 300bp, two of them within 120bp. However, in the orientation that requires 66bp of 3' flanking sequence, no homologies to 3' conserved sequence are found within 400bp of the insertion site. The absence of any 3' conserved sequence in the 66bp of yeast sequence suggests however, that other DNA sequences may be important as determinants of ARS efficiency.

A detailed mutational analysis of the C2G1 ARS which will allow an assessment of the biological significance of the 3' conserved sequence is now in progress.

ACKNOWLEDGEMENTS

We thank P. Pizzigallo for technical assistance and Drs. W. Fangman, M. Newlon and L. Ripley for helpful comments on the manuscript. This work was supported by NIH Grants GM21510 and GM 35679 to CSN, SERC grant GR/C34656 to SGO and R.W. Davies, and by a NATO collaborative award to CSN and SGO. TGP was supported by NIH Predoctoral Genetics Training Grant GM07091 awarded to the University of Iowa.

REFERENCES

1. Williamson, D.H. (1985) Yeast 1,1-14.
2. Struhl, K., Stinchcomb, D.T., Sherer, S. and Davis, R.W. (1979) Proc. Nat. Acad. Sci., USA 76,1035-1039.
3. Stinchcomb, D.T., Struhl, K., and Davis, R.W. (1979) Nature 282,39-43.
4. Kingsman, A.J., Clarke, L., Mortimer, R.K. and Carbon, J. (1979) Gene 7,141-152.
5. Zakian, V.A. and Scott, J.F. (1982) Molec. Cell. Biol. 2,221-232.
6. Fangman, W.L., Hice, R.H. and Chlebowicz-Sledziewska, E. (1983) Cell 32,831-838.
7. Beach, D., Piper, M. and Shall, S. (1980) Nature 284,185-187.
8. Chan, C.S.M. and Tye, B.-K. (1980) Proc. Nat. Acad. Sci., USA 77,6329-6333.
9. Newlon, C.S. and Burke, W.G. (1980) ICN-UCLA Symp. on Molec. and Cell. Biol. 19,399-409.
10. Stinchcomb, D.T., Mann, C., Selker, E., and Davis, R.W. (1981) ICN-UCLA Symp. on Cell. and Molec. Biol. 22,473-488.
11. Broach, J.R., Li, Y.-Y., Feldman, J., Jayaram, M., Abraham, J., Nasmyth, K., and Hicks, J.B. (1983) Cold Spring Harbor Symp. Quant. Biol. 47,1165-1173.
12. Celniker, S.E., Sweder, K., Srienc, F., Bailey, J.E. and Campbell, J.L. (1984) Molec. Cell. Biol. 4,2455-2466.
13. Kearsey, S. (1984) Cell 37,299-307.
14. Bouton, A.H. and Smith, M.M. (1986) Molec. Cell. Biol. 6, 2354-2363.
15. Srienc, F., Bailey, J.F. and Campbell, J.L. (1985) Molec. Cell. Biol. 5,1676-1684.
16. Strathern, J.N., Newlon, C.S., Herskowitz, I., and Hicks, J.B. (1979) Cell 18,309-319.
17. Newlon, C.S., Green, R.P., Hardeman, K.J., Kim, K.E., Lipchitz, L.R., Palzkill, T.G., Synn, S., and Woody, S.T. (1985) UCLA Symp. on Molec. Cell. Biol. 33, in press.
18. Devenish, R.J. and Newlon, C.S. (1982) Gene 18,277-288.
19. Miller, J.H. (1972) Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
20. Stinchcomb, D.T., Thomas, M., Kelly, J., Selker, E., and Davis, R.W. (1980) Proc. Nat. Acad. Sci., USA 77,4559-4563.
21. Maniatis, T., Fritsch, E.F., and J. Sambrook (1982) Molecular Cloning A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
22. Garger, S.J., Griffith, O.M. and Grill, L.K. (1983) Biochem. Biophys. Res. Comm. 117,835-842.
23. Guyer, M.S. (1983) Meth. in Enzymol. 101,362-369.
24. Guyer, M.S. (1978) J. Mol. Biol. 126,347-365.
25. Maxam, A.M. and Gilbert, W. (1980) Meth. Enzymol. 65,499-560.
26. Sanger, F., Nicklen, S. and Coulson, A.R. (1979) Proc. Nat. Acad. Sci., USA 74,5463-5467.
27. Deininger, P.L. (1983) Analyt. Biochem. 129,216-223.
28. Lagrimini, L.M., Brentano, S.T. and Donelson, J.E. (1984) Nuc. Acids Res. 12,605-614.
29. Staden, R. (1977) Nuc. Acids. Res. 4,4037-4051.
30. Staden, R. (1978) Nuc. Acids. Res. 3,1013-1015.
31. Feldman, H., Olah, J., and Friedenreich, H. (1981) Nuc. Acids Res. 9,2949-2959.
32. Skryabin, K.G., Eldarov, M.A., Larionov, V.L., Bayev, A.A., Klootwijk, J., de Regt, V.C.H.F., Vildman, G.M., Planta, R.J., Gerogiev, O.I., and Hadjiolou, A.A. (1984) Nuc. Acids Res. 12,2955.
33. Tschumper, G. and Carbon J. (1980) Gene 10,157-166.

34. Tschumper, G. and Carbon, J. (1982) J. Mol. Biol. 156,293-307.
35. Broach, J.R. and Hicks, J.B. (1980) Cell 21,501-508.
36. Hartley, J. and Donelson, J.E. (1980) Nature 286,860-865.
37. Oka, A., Sasaki, H., Sugimoto, K. and Takanami, M. (1984) J. Mol. Biol. 176,443-458.
38. McClure, W.R. (1985) Ann. Rev. Biochem. 54,171-204.
39. Hall, B.D., Clarkson, S.G. and Tocchini-Valentine (1982) Cell 29,3-5.
40. Nagawa, F. and Fink, G.R. (1985) Proc. Natl. Acad. Sci. USA 82,8557-8561.
41. Hahn, S., Hoar, E.T. and Guarante, L. (1985) Proc. Natl. Acad. Sci. USA 82,8562-8566.
42. Fangman, W.F., McCarroll, R.M., Brewer, B.J., and Zweifel, S.J. (1986) J. Cell. Biochem., Suppl. 10B, 153.
43. Smith, M.M. and Murray (1983) J. Mol. Biol. 169,641-661.
44. Bergsma, D.J., Olive, D.M., Hartzell, S.W. and Subramanian, K.N. (1982) Proc. Natl. Acad. Sci. USA 79,381-385.
45. DeLucia, A., Deb, S., Parten, K. and Tegtmeyer, P. (1986) J. Virology 57,138-144.