Volume 12 Number 2 1984 Nucleic Acids Research

Characterization of human chromosomal DNA sequences which replicate autonomously in *Saccharomyces cerevisiae*

J.F.Montiel, C.J.Norbury, M.F.Tuite, M.J.Dobson, J.S.Mills, A.J.Kingsman and S.M.Kingsman*

Department of Biochemistry, South Parks Road, Oxford OX1 3QU, UK

Received 4 October 1983; Revised and Accepted 28 November 1983

ABSTRACT

We have characterised two restriction fragments, isolated from a "shotgun" collection of human DNA, which function as autonomously replicating sequences (ARSs) in Saccharomyces cerevisiae. Functional domains of these fragments have been defined by subcloning and exonuclease (BAL 31) deletion analysis. Both fragments contain two spatially distinct domains. One is essential for high frequency transformation and is termed the Replication Sequence (RS) domain, the other, termed the Replication Enhancer (RE) domain,has no inherent replication competence but is essential for ensuring maximum function of the RS domain. The nucleotide sequence of these domains reveals several conserved sequences one of which is strikingly similar to the yeast ARS consensus sequence.

INTRODUCTION

Little is known about the structural organisation and regulation of eukaryotic DNA replication (reviewed in 1 and 2). It is still not clear whether there are specific origins of replication. The non-random spacing of initiation sites in Drosophila embryos and in somatic cells (3) indicates that this might be the case. However initiation of DNA replication in Xenopus laevis eggs does not require specific sequences (4). There is also evidence for functional differences between initiation sites implying some element of specificity in the replication mechanism. For example, in mammalian cells the same regions of DNA are replicated at the same times during successive S phases (5) and different initiation points may be used in the same cells under different environmental conditions (6). The idea that eukaryotic DNA replication may involve specific sequences was strengthened by the discovery that a specific subset of restriction fragments from yeast chromosomal DNA can replicate in yeast independently of the chromosome (7). These autonomously replicating sequences (ARSs) increase the transforming ability of yeast integrative plasmids (eg.8) by 10^3 to 10^5 fold (9,10) and this property has provided the basis for

© IRL Press Limited, Oxford, England.

selection of ARSs from many eukaryotic genomes. ARSs have been identified in DNA obtained from yeast (9-15), Neurospora crassa, Dictyostelium discoideum, Ceanorabditis elegans, Drosophila melanogaster and Zea mays (13), Tetrahymena (16), Xenopus laevis mitochondria (17), X.laevis chromosomes (M.J.Dobson, A.J Kingsman and S.M. Kingsman, unpublished data; S.Kearsey, personal communication) and as reported here, human chromosomes. No ARSs have be isolated from bacterial chromosomal DNA (13). The frequency of isolation of ARSs from yeast chromosomal DNA closely reflects the frequency of replication origins determined by electron microscopic techniques (18). Furthermore DNA replication initiates specifically in the region of yeast ARSs in vitro and specific protein complexes involved in replication also bind specifically to ARSs in vitro (19,20). In addition different yeast ARSs replicate at different but specific times during the yeast cell cycle (21). These studies suggest that yeast chromosomal sequences with ARS activity may contain biologically significant sequences which are important for the initiation and/or regulation of eukaryotic DNA replication. It is possible that sequences isolated from other eukaryotes which have ARS activity in yeast might also be involved in DNA replication in the homologous systems. Sequence comparisons of several yeast ARSs (14,15,22-26) has revealed an 11 bp consensus sequence, 5'-TAAAPyAPyAAPuA-3' first noted by Stinchcomb et al (23). However not all yeast ARSs contain the above sequences (27) and they are not found in a region of Tetrahymena rDNA which has ARS activity (16). Other possibly significant features such as TATA boxes, GC rich regions and other homologies derived from pairwise comparisons of yeast ARSs have also been noted (26). This paper presents a detailed examination of the extent of structural and functional conservation between yeast ARSs and two ARSs isolated from human chromosomal DNA.

MATERIALS AND METHODS

Bacterial and yeast strains and media

E.coli strain Sf8 = C600, hsdRk, hsdMk, lop11, leu B6 recBC. *Saccharomyces cerevisiae* strain MD40-4c = α, ura2, leu2-3, leu2-112, his3-11, his3-15, trp1. *E.coli* were grown in Luria broth (28). Yeast media were prepared according to Hawthorne and Mortimer (29).

Enzymes

Restriction endonucleases and ligase were purchased from Bethesda Research Laboratories (BRL) and were used according to the suppliers

instructions.

Yeast transformation

The method described by Hinnen et al (8) was used

Analysis of yeast transformants

Transformants were tested for ploidy by measuring the frequency of mutation to canavanine resistance (9). Haploid transformants were grown for 24 hr under selection and used to inoculate fresh selective media at 10^4 cells/ml. Growth was measured by cell counts in a haemocytometer. Plasmid stability was determined by growing the cells in selective media and then transferring to non-selective media for 4hr. The number of cells containing plasmid after growth in selective and non-selective media was determined by washing the cells in water and then plating onto duplicate selective and non-selective plates and the percent loss of plasmid per generation calculated.

DNA

Plasmid DNA was isolated from E.coli preparatively as described by Chinault and Carbon (30) and for rapid analysis by the method of Holmes and Quigley (31). Total yeast DNA was isolated according to Cryer et al (32). Plasmids were rescued from yeast transformants by the method of Ferguson et al (33). Hela cell DNA isolated as described by Cook and Brazell (34) was a gift from Dr. P.R.Cook,Sir William Dunn School of Pathology, Oxford. Human placental DNA was a gift from Dr. I.W. Craig, Department of Genetics, Oxford.

Southern transfers, hybridisations and in vitro labelling

Restriction endonuclease digested DNAs were fractionated by agarose gel electrophoresis and the fragments were transferred to nitrocellulose by Southerns procedure (35). Plasmids were labelled by nick translation (36) using ^{32}P-TTP (Amersham International). Hybridisations were carried out in 0.3M NaCl, 0.03M sodium citrate, 0.02% ficoll, 0.02% BSA, 0.02% polyvinylpyrrolidone at 65^{o}C for 48 hrs.

Deletion formation and linker ligation

Plasmids were cleaved with appropriate restriction enzymes and digested with BAL 31 exonuclease (BRL) as described previously (37). Synthetic Bgl II linkers (5'-CAAAAGATCTTTTG-3') and synthetic Eco RI linkers (5'-GGAATTCC-3') were generously provided by Dr. M.A.W. Eaton, Celltech Ltd. Linkers were phosphorylated and blunt end ligations performed as described previously (37).

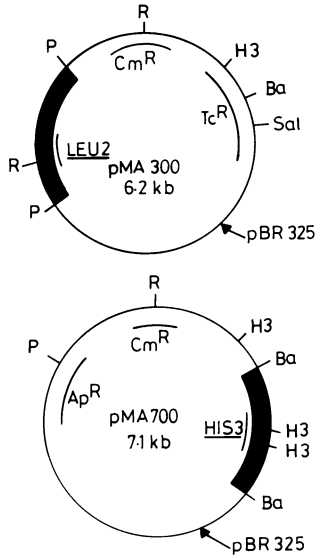


Figure 1. Yeast integrative vectors. Thin line = pBR325. Thick line = yeast chromosomal DNA. Ba = Bam HI; H3 = Hind III; P = Pst I; Sal = Sal I; R = Eco RI; Ap^R, Cm^R and Tc^R are respectively, resistance to ampicillin, chloramphenicol and tetracycline.

DNA sequencing

The nucleotide sequence of human ARS fragments was determined by the dideoxy chain termination method of Sanger et al (38). Single stranded templates were isolated by subcloning either Hind III-Bgl II or Eco RI-Bgl II fragments into M13mp8 or M13mp9 (39).

RESULTS

Isolation of human DNA fragments with ARS activity

We have used three plasmids for the isolation and characterization of ARSs; pGT6 , which contains a 2 kb Pst I yeast chromosomal LEU2 fragment inserted at the Pst I site in pBR322 (22); pMA300 which contains the same 2 kb LEU2 as pGT6 but inserted into pBR325 (Figure 1) and pMA700 which contains a 1.85 kb yeast HIS3 fragment (40) both inserted into pBR325 (41) (Figure 1). These plasmids rely upon integration into the yeast chromosome for maintenance and consequently transform yeast at a very low frequency (about 2 transformants per μg of vector DNA). Plasmids capable of autonomous replication transform yeast at a 10^3-10^5 fold higher frequency. This difference in transformation frequency is used as a selection system

for ARS elements. If the frequency of transformation of an integrative vector is increased above 2/µg by the insertion of random DNA fragments then it is likely that those fragments have ARS activity. Human placental DNA and Hela cell DNA were cleaved to completion with Hind III and Eco RI respectively and ligated to appropriately cleaved pGT6 or pMA700. The ligation mixes (2µg of vector and 5µg of human DNA) were used to transform yeast to either leucine or histidine independence. A total of 25 transformants was obtained, 18 from the pGT6 reaction and 7 from the pMA700 reaction. In parallel transformations with pGT6 and pMA700 alone no transformants were obtained. Upon subculturing the transformants on the same medium on which they had been selected, one transformant grew as well as MD40-4c and probably had pMA700 integrated into the yeast genome. Of the other 24 transformants, 16 failed to grow and 8 grew but showed relatively few colonies compared with the number plated. The 16 that failed to grow upon subculture presumably contained autonomous molecules so unstable that they were lost completely during the establishment of the primary transformant colony. The 8 transformants that grew poorly must maintain their plasmids more efficiently. Four of the 8 unstable transformants were chosen for further study. Three transformants designated T52, T53 and T54 contained Eco RI fragments of Hela cell DNA inserted into pMA700 and the other designated T50 contained a Hind III fragment of placental DNA inserted into pGT6. All four transformants were haploid. To rescue plasmid DNA from these transformants, DNA enriched for small circular supercoiled molecules was isolated and used to transform E.coli to leucine independence in the case of DNA from T50 or chloramphenicol resistance in the case of DNA from T52, T53 and T54. Plasmid DNA was prepared from an E.coli transformant obtained with each of the four DNA preparations. Restriction enzyme analysis showed that the plasmid isolated from T50, designated pMA50, contained a 4.2 kb Hind III fragment in pGT6 and the plasmids isolated from T52, T53 and T54, designated pMA52, pMA53 and pMA54 contained 2.6 kb, 6.65 kb and 4.45 kb Eco RI fragments in pMA700. Restriction maps of the human DNA inserts in these plasmids are shown in figure 2 and they exclude the possibility that they are derived from human mitochondrial DNA (42).

Location of the human ARSs to human chromosomal DNA

Since the selection system for ARS elements in yeast is powerful it was important to establish that the sequences in pMA50, pMA52, pMA53 and pMA54 were in fact derived from human DNA. Hela cell DNA was cleaved with

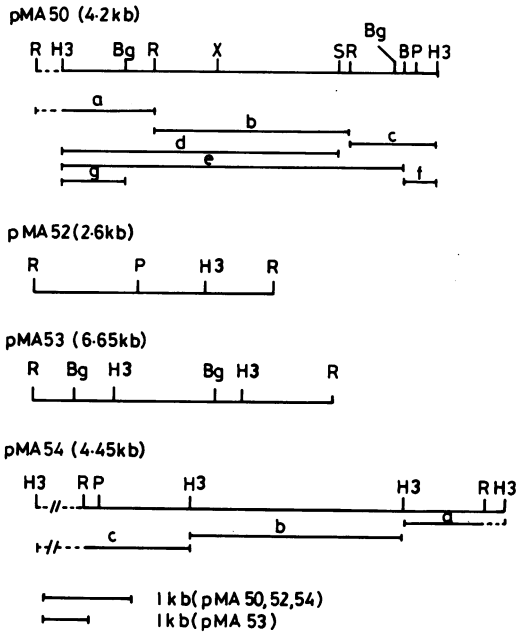


Figure 2. Partial restriction maps of human ARS containing fragments and subcloning strategy. Solid line = human DNA ; Dotted line = vector DNA. Bars below pMA50 and pMA54 show the fragments which were subcloned to localise the ARS (see text). B = Bam HI; Bg = Bgl II; H3 = Hind III; P = Pst I; S = Sal I; R = Eco RI; X = Xho I.

Eco RI and probed with each of the ^{32}P labelled plasmids in a Southern blot analysis (Data not presented). In each case fragments from Hela cell DNA hybridised with the probe and the sizes of these fragments were consistent with the maps in figure 2. These data confirm the human origin of the inserts in the four plasmids and also show that the human ARS elements are unique and/or too short to cross hybridize.

The human DNA fragments replicate autonomously in yeast

There are two major diagnostic features of ARS containing plasmids. The first is that they transform yeast at a relatively high frequency and the second is that they are unstable. Our ability to rescue plasmid DNA from the transformants is consistent with autonomous replication and preliminary observation of the transformants suggested that they contained unstable plasmids. To confirm the autonomy of the four plasmids and to quantitate their behaviour, the transformation efficiencies and stabilities were compared with the archetypal yeast ARS containing plasmid YRp7 (9,10)

Table 1. Properties of yeast strains transformed with ARS plasmids

Plasmid	Transformation frequency (colonies/μg)	Stability (per cent plasmid loss per generation)
YRp7	3.5×10^4	22
pMA50	6.1×10^4	36
pMA52	3.2×10^4	43
pMA53	4.0×10^3	>50
pMA54	1.0×10^4	45

(table 1). The human ARS plasmids are able to transform yeast at frequencies comparable to YRp7 and they are all unstable under selective and non-selective conditions. In all cases the human ARS plasmids were apparently less stable than YRp7. In addition when total undigested DNA from yeast transformants T50, T52, T53 and T54 is probed with ^{32}P labelled pBR322 in a Southern blot analysis bands are obtained which comigrate with the supercoiled pure plasmid (data not presented). Together these data prove that plasmids pMA50, pMA52, pMA53 and pMA54 contain human DNA sequences that are capable of autonomous replication in yeast.

Localization of the ARS activity in two of the human DNA fragments

Two of the human DNA fragments were chosen for a detailed analysis; these were pMA50 which contains a 4.2 kb Hind III fragment from human placental DNA and pMA54 which contains a 4.45 kb Eco RI fragment from Hela cell DNA. To localize the ARS activity on these plasmids various restriction fragments, indicated in figure 2, were subcloned into either pMA700 or pMA300 as appropriate. The ability of each subclone to transform yeast at a high frequency was tested. In the case of subclones from pMA50, fragment g was sufficient to give high frequency transformation and fragments which did not encompass this region were not capable of high frequency transformation. The human ARS activity had therefore been localized to a 0.7 kb Hind III-Bgl II fragment. In the case of pMA54 the ARS activity was localised to the 2.3 kb Hind III fragment b. The Hind III-Bgl II fragment g from pMA50 replaces the small Hind III-Bam HI fragment in pMA300 to give plasmid pMA501 and the insert is referred to as human ARS1. The Hind III fragment b from pMA54 is inserted at the Hind III site in pMA300 to give plasmid pMA541 and the insert is referred to as human ARS2.

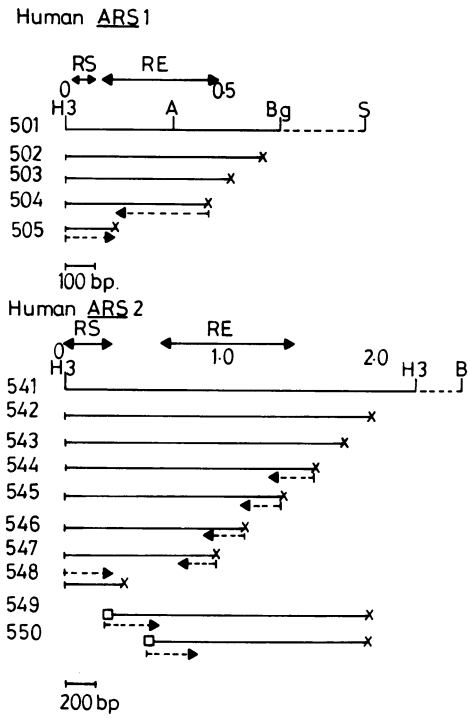


Figure 3. Deletion derivatives of human ARS1 and human ARS2. Solid lines = Human DNA; Dotted lines = vector DNA. Numbers above the maps indicate coordinates in kb. Dotted arrows show the region and direction of sequencing of the various deleted fragments. A = Acc I; B-Bam HI, H3 = Hind III; S = Sal I; x = synthetic Bgl II linker; ◘ = synthetic Eco RI linker. The Replication Sequence (RS) and Replication Enhancer (RE) domains are indicated by arrows above the maps.

In order to localize the ARS activity for sequencing a series of deletions extending into the right side of the fragments, as drawn in figure 3, was constructed using limited digestion with the exonuclease BAL 31. pMA501 was cleaved at the Sal I site and pMA541 was cleaved at the Bam HI site in the vector DNA and the plasmids were digested with BAL 31 for appropriate times. After digestion the reaction products were blunt end ligated with synthetic Bgl II linkers and the ligation mixes used to transform E.coli to chloramphenicol resistance. Rapid DNA preparations from transformants were analysed by digestion with Hind III and Bgl II to identify molecules which contained the linker and to determine the size of the deleted human fragments. A second set of deletions of human ARS2 starting at the left hand end was made. As there are no unique restriction sites near the left

Table 2. Properties of deletion derivatives of human ARS plasmids

Plasmid	Human DNA (kb)	Transformation frequency (Colonies/μg $\times 10^{-4}$)	Colony size[a]
YRp7	-	1.0	L
pMA501	0.7	0.9	L
pMA502	0.65	0.3	L
pMA503	0.55	0.3	L
pMA504	0.46	0.8	L
pMA505	0.13	0.6	S
pMA542	2.00	2.9	L
pMA543	1.80	1.5	L
pMA544	1.60	1.7	L
pMA545	1.47	0.5	L
pMA546	1.20	0.5	S
pMA547	1.07	0.8	S
pMA548	0.52	0.7	S
pMA549	1.70	0	-
pMA550	1.40	0	-

a. Colony size was judged subjectively four days after transformation. L = large, S = small.

hand end of the fragment in pMA541, the plasmid pMA542 from the first deletion series (figure 3) was used. This plasmid was cleaved at the unique Hind III site and after BAL 31 deletion the reaction products were ligated with synthetic Eco RI linkers and used to transform E.coli to chloramphenicol resistance. Pure DNA of selected plasmids from these deletion series was prepared and the size and location of the different deleted fragments are shown in figure 3. Each of the deleted plasmids was tested for the ability to transform yeast at a high frequency (table 2). The deletion analysis indicated that in the case of human ARS1 the human DNA insert could be reduced to a 131 bp fragment in pMA505 which still retained the ability for high frequency transformation. Similarly human ARS2 could be deleted from the right hand side down to a 522 bp fragment in pMA548 which still showed high frequency transformation. Deletions from the left hand side of human ARS2 produced plasmids pMA549 and pMA550 which had lost 308 bp and 683 bp respectively and were completely unable to transform yeast at a high frequency. This result indicated that the sequence responsible for high frequency transformation in human ARS2 resided within 308 bp at the extreme left of the fragment as drawn in figure 3. Although pMA505 and pMA548 both contained small DNA fragments

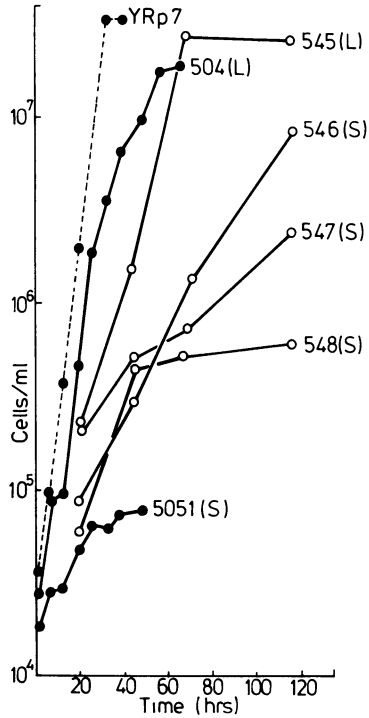


Figure 4. Growth curves of yeast transformed with various human ARS plasmids. Transformants were grown to mid log phase in selective media then inoculated at a density of about 10^4 cells/ml into fresh selective media. Colony size of the transformants on solid media is indicated in parentheses, L = large, S = small. pMA5051 is the same as pMA505 except that the insert has been recloned into pMA300 (see text).

which were capable of high frequency transformation both plasmids produced transformants which were very slow growing. This was observed initially as the production of very small transformants which took four days to achieve a size large enough for subculture and was confirmed by following growth under selective conditions in liquid media. The growth curves for some of the transformants are shown in figure 4. All the transformants relying on human DNA sequences for plasmid replication have longer doubling times than YRp7 transformants (10), of the order of 6-10hrs. These results and the data in table 1 indicate that the human DNA sequences are not as replication competent as yeast sequences. The maximum cell density achieved by pMA548 after 120 hrs growth is only $6x10^5$cells/ml whereas transformants containing plasmids with larger inserts such as pMA504 and pMA545 reach

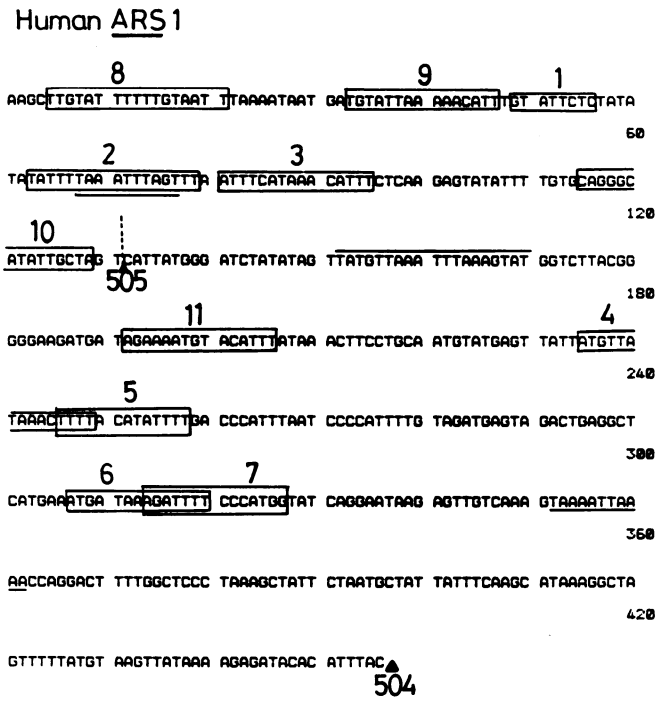
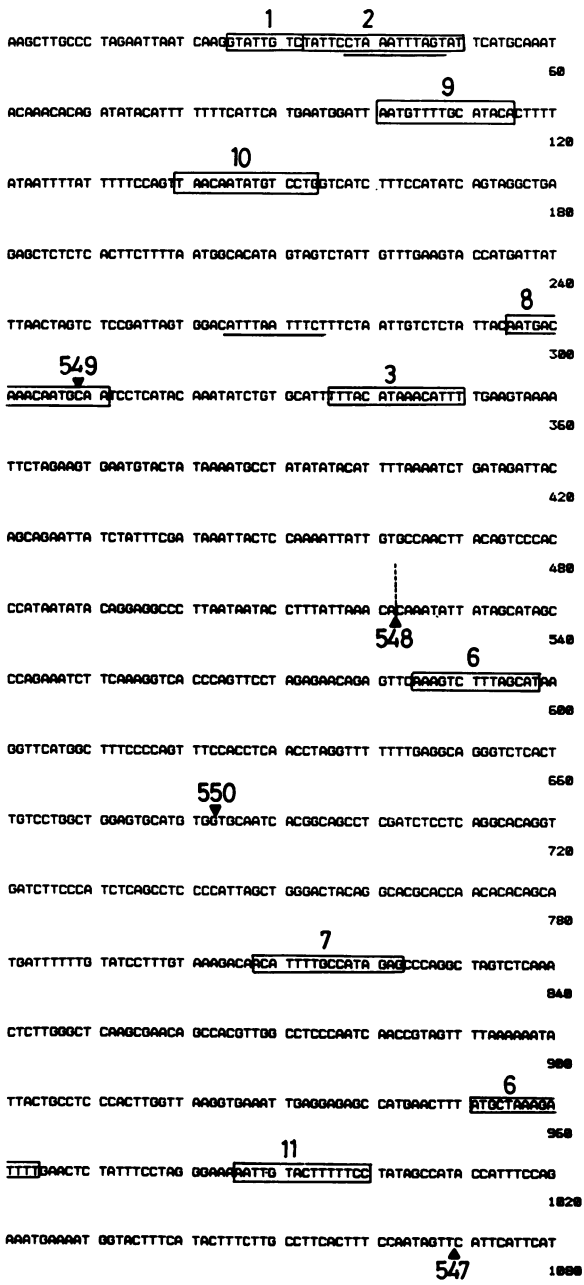


Figure 5. Nucleotide sequence of the RS and RE regions of Human ARS1. The sequence of one strand 5'-3' is shown. Regions of homology shared with human ARS2 are boxed. Underlined sequences share homology with the yeast ARS consensus sequence (23). The RS domain lies to the left of the dotted line at position 131, the RE domain lies to the right of this line. Numbers with arrows indicate the position of deletion end points.

$2x10^7$ cells/ml after about 45 hrs. Some of the deleted plasmids, for example pMA547 and pMA546, had intermediate growth characteristics. Similar results were obtained with the pMA501 series (data not shown). This result suggested that while the sequences present in pMA505 and pMA548 were sufficient and necessary for high frequency transformation there may be additional spatially distinct sequences which are required to ensure maximum replication competence and/or to increase mitotic stability. As BAL 31 deletion is bidirectional the progressive loss of human DNA sequences is accompanied by a loss of vector sequences. In order to confirm that the reduction in replication efficiency was in fact due to the loss of human DNA sequences the human DNA insert from pMA505 was recloned into pMA300 to produce plasmid pMA5051. This plasmid transformed yeast at a high frequency but again grew extremely slowly in selective media (figure 4) indicating

Human ARS 2


1
2
AAGCTTGCCC TAGAATTAAT CAAGGTATTG TCTATTCCTA AATTTAGTAT TCATGCAAAT
60
9
ACAAACACAG ATATACATTT TTTTCATTCA TGAATGGATT AATGTTTTGC ATACACTTTT
120
10
ATAATTTTAT TTTTCCAGTT AACAATATGT CCTGGTCATC TTTCCATATC AGTAGGCTGA
180
GAGCTCTCTC ACTTCTTTTA ATGGCACATA GTAGTCTATT GTTTGAAGTA CCATGATTAT
240
8
TTAACTAGTC TCCGATTAGT GGACATTTAA TTTCTTTCTA ATTGTCTCTA TTACAATGAC
300
549
3
AAACAATGCA ATCCTCATAC AAATATCTGT GCATTTTTAC ATAAACATTT TGAAGTAAAA
360
TTCTAGAAGT GAATGTACTA TAAAATGCCT ATATATACAT TTTAAAATCT GATAGATTAC
420
AGCAGAATTA TCTATTTCGA TAAATTACTC CAAAATTATT GTGCCAACTT ACAGTCCCAC
480
CCATAATATA CAGGAGGCCC TTAATAATAC CTTTATTAAA CACAAATATT ATAGCATAGC
548
540
6
CCAGAAATCT TCAAAGGTCA CCCAGTTCCT AGAGAACAGA GTTCAAAGTC TTTAGCATAA
600
GGTTCATGGC TTTCCCCAGT TTCCACCTCA ACCTAGGTTT TTTTGAGGCA GGGTCTCACT
660
550
TGTCCTGGCT GGAGTGCATG TGGTGCAATC ACGGCAGCCT CGATCTCCTC AGGCACAGGT
720
GATCTTCCCA TCTCAGCCTC CCCATTAGCT GGGACTACAG GCACGCACCA ACACACAGCA
780
7
TGATTTTTTG TATCCTTTGT AAAGACACCA TTTTGCCATA GAGCCCAGGC TAGTCTCAAA
840
CTCTTGGGCT CAAGCGAACA GCCACGTTGG CCTCCCAATC AACCGTAGTT TTAAAAAATA
900
6
TTACTGCCTC CCACTTGGTT AAGGTGAAAT TGAGGAGAGC CATGAACTTT ATGCTAAAGA
960
11
TTTTGAACTC TATTTCCTAG GGAAAAATTG TACTTTTTCC TATAGCCATA CCATTTCCAG
1020
AAATGAAAAT GGTACTTTCA TACTTTCTTG CCTTCACTTT CCAATAGTTC ATTCATTCAT
547
1080

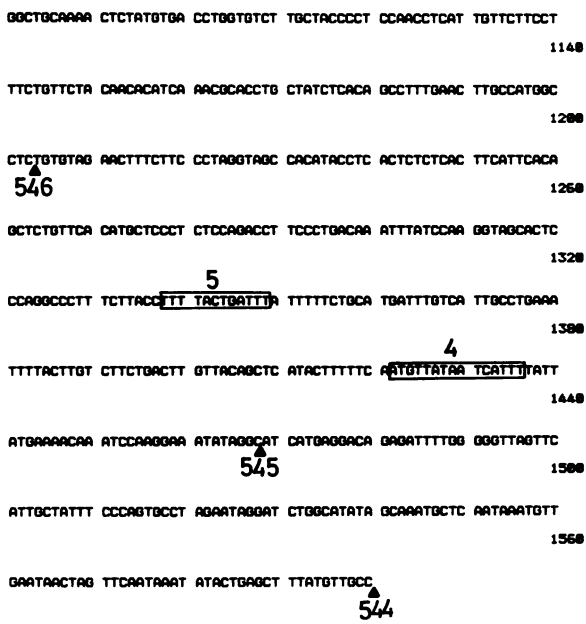


Figure 6. Nucleotide sequence of the RS and RE domains of human ARS2. Legend as for figure 6 except that the RS region lies to the left of the dotted line at position 522 and the RE lies to the right of this line. End points of deletions made from the left end of the fragment are indicated above the sequence and deletions from the right of the fragment are indicated below the sequence

that vector sequences could not restore replication efficiency. We have designated the region responsible for high frequency transformation which is present on pMA505 and pMA548 as the Replication Sequence (RS) domain and the region which is responsible for full replication competence as the Replication Enhancer (RE) domain. The RS domain in human ARS1 is present on pMA505 and the RE domain is present on pMA504. The RS domain in human ARS2 is present on pMA548, however the RE domain is difficult to locate, it must be present on plasmids pMA541-pMA545 but plasmids pMA546 and pMA547 appear to have intermediate replication properties (figure 4) and the region could span at least 1.0 kb.

Nucleotide sequence of the RS and RE domains of human ARS1 and ARS2

The sequencing strategy is outlined in figure 3. The nucleotide sequences of the RS and RE domains of human ARS1 were obtained by sequencing the Hind III- Bgl II fragment from pMA505 and the overlapping Bgl II-Hind III fragment from pMA504. The RS domain of human ARS2 was

sequenced using the Hind III-Bgl II fragment from pMA548 and the overlapping Eco RI-Bgl II fragment from pMA549. The sequence of RE domain of human ARS2 was obtained from sequencing the overlapping Bgl II-Hind III fragments (figure 3.) The sequence of the RS and RE domains of human ARS1 and ARS2 are shown for one strand in figures 5 and 6 respectively. The limits of the RS domain in human ARS1 are between nucleotide 1 and 131 which is the deletion end point in pMA505. The limits of the RS domain in human ARS2 are between nucleotide 1 and 308 as the loss of these nucleotides in pMA549 renders the plasmid incapable of high frequency transformation. The RE domain of human ARS2 may lie anywhere within the remaining 1292 nucleotides. Both strands of each sequence (1-456 bp for human ARS1 and 1-1600 bp for human ARS2) were compared using the Oxford HOMOL 2 homology search programme (A.J. Kingsman, unpublished data). Many short stretches, less than 10 bp, of homology were observed but these will not be analysed in the present paper. There are also cases where a sequence was found in the RS domain of one human ARS and in the RE domain of the other; such sequences in isolation cannot be significant in terms of RS and RE function and are not analysed in this paper. The real biological significance of any homologies can only be assessed by detailed functional analysis of mutants but we have indicated some potentially interesting regions in figures 5 and 6. There are six stretches of homology unique to the RS domain of each human ARS, 1-3 are derived from comparing the strands shown in figures 5 and 6 and 8-10 from comparing the human ARS1 strand shown in figure 5 with the opposite strand of human ARS2 (not shown). Homology block 1 is small and only considered significant because of its similar spatial relationship to blocks 2 and 3 in each ARS. Homology block 3 cannot be a significant replication sequence per se because it is present on pMA549 which is not replication competent. However as it is present on pMA548 and pMA505 a potential involvement in determining replication competence cannot be excluded. The longest homology block, 2, shows 15/18 conserved positions and gives a "consensus" sequence of 5′-TATTPy(Py)TAAATTTAGTA/TT-3′. A sequence related to this homology block is also found in the RE domain of ARS1 (fig. 5) and although this is not required for the ARS phenotype a role in determining maximum efficiency of autonomous replication cannot be excluded. Homology block 2 shows some similarities with sequences thought to be significant in yeast ARSs (23) and the human sequences are compared with several yeast ARSs in figure 7. There are additional short sequences located in the RE domains of both

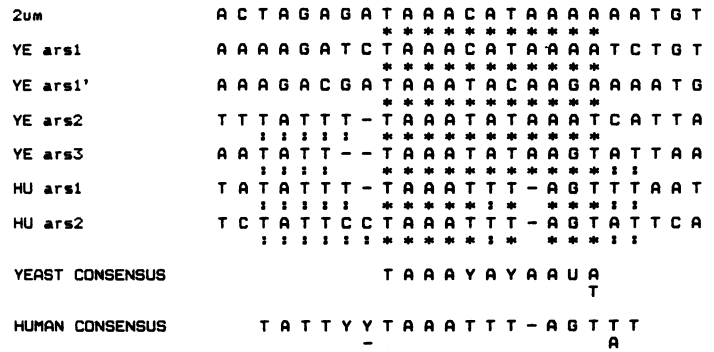


Figure 7. Comparison of the typeI RS in human and yeast ARSs - = space left to allow aligning of the sequences. * = homology with the yeast consensus sequence : = homology with the human consensus sequence. Y = pyrimidine; U = purine. Sequence information for yeast ARS1, ARS1', ARS2, ARS3 and 2 μm are taken from references 22,26,23 and 14.

human ARSs which share homology with the yeast ARS consensus sequence (figs. 5 and 6) but these cannot function as replication sequences. There are five homology blocks in the RE domains, numbers 4-7 are derived from comparing the strands shown and 11 is derived from comparing the strand shown for human ARS1 with the opposite strand of human ARS2 (not shown) Block 6 is repeated in human ARS2 on opposite strands. It is perhaps interesting to note that blocks 4 and 5 and 6 and 7 overlap and are clustered in human ARS1 but they are dispersed in human ARS2.

DISCUSSION

We have demonstrated for the first time that sequences can be isolated from human chromosomal DNA which will replicate autonomously in yeast. This result extends previous observations about the functional conservation of ARSs within eukaryotes (13). Such conservation across a broad range of eukaryotes implies that ARSs may play a fundamental role in some aspect of eukaryotic DNA replication. The restriction maps of the human ARSs analysed here show that they are clearly derived from chromosomal and not mitochondrial DNA. They are also not associated with any repetitive sequences and do not show any cross hybridisation under stringent conditions. The frequency of isolation of ARSs was low and recovery of primary transformants after replating on the same medium was less than 50%. The four human ARS transformants selected for further analysis grew more

slowly and were less stable than yeast transformed with YRp7 (table 1 and figure 4). These results indicate that although ARSs can be isolated from human DNA they appear to function inefficiently in the heterologous environment. It is likely that many potential ARSs may have been undetected in the original transformation because of this inefficient function and therefore only a subset of human ARSs may have been studied.

The deletion analysis to generate small fragments which still retained ARS activity has indicated that the gross structure of human ARSs may be complex. We have identified two spatially distinct domains which are required to give high frequency transformation and to maximize efficient plasmid replication and/or maintenance. The RS domain must contain one or more sequences which are required for high frequency transformation; plasmids which lack this domain have no ARS activity. Plasmids which carry only the RS domain transform yeast at high frequency but the resulting transformants grow very poorly (figure 4). The second domain, RE, is required to augment the function of the RS domain to produce vigourously growing transformants. Plasmids which carry only the RE domain such as pMA549 never transform yeast at high frequency i.e. they do not behave like ARS plasmids. The RE domain must contain one or more sequences which interact in some way with the RS domain. It is likely that several sequences are involved because the RE domain in human ARS2 spans at least 1 kb. This bipartite structure of human ARSs is similar to that proposed for yeast ARSs by Stinchcomb et al (23) who showed that a 626 bp fragment containing the ARS1 consensus sequences is sufficient to give high frequency transformation but the transformants have a doubling time of 12 hrs. An adjacent 212 bp fragment was required to produce vigourously growing transformants. Similarly the yeast 2 μm replication origin contains a core 75 bp which is probably sufficient for high frequency transformation but additional flanking sequences are required for maximum replication efficiency (14). Similar results have also been obtained in an analysis of an ARS isolated from Candida utilis mitochondrial DNA (43). One of the sequences implicated as important in the RS domain of yeast ARSs is the consensus sequence 5'-T/AAAAPyAPyAAPuA/T-3' (23). The human ARSs share several conserved sequences in the RS domain. It may be significant that one of these sequences is very similar to several yeast RSs and contains a sequence that is related to the yeast consensus sequence (fig. 7). The fact that the majority of ARSs sequenced to date contain a sequence related to this yeast consensus suggests either that this is the major sequence

involved in eukaryotic DNA replication or simply that it is the sequence which functions most efficiently in yeast. We refer to this group of related replication sequences as type I replication sequences. Clearly sequences other than the type I sequences can have ARS activity as they are not found on at least one yeast ARS (27) and a Tetrahymena ARS (16) although the latter may be a specialized replicon. It is also possible that there are other important sequences and the additional conserved sequences in the human RS domains may be functional. It seems unlikely that the minimal 11 bp type I replication sequence is sufficient for replication to occur. There is a second type I sequence on both yeast ARS1 and 2 μm and no replication eyes or protein complexes are found associated with these (20). There are also type I sequences at the MAT **a** and MAT α loci but neither of these sequences have ARS activity (14). Furthermore short sequences similar to the yeast ARS consensus sequence are found outside the RS domains of both human ARSs (figs. 5 and 6). All these observations suggest that if the type I RS is functionally significant then the precise nucleotide sequence environment is critical to allow the function to be expressed. Not all ARSs may require either additional sequences in the RS domain or an RE domain. It has been shown that as little as 57 bp is sufficient for full replication competence of the ARS associated with the yeast HO gene implying that any sequences required in addition to the type I sequences must be small and closely associated with the type I sequence (15). The RS domains of all ARSs may contain sequences that interact with replication enzymes and/or define the site of initiation of DNA replication as has been shown for yeast type I replication sequences (19,20). Other specific sequences may be required to control the activity of the replication origin, for example in determining the time of replication during the cell cycle(21). Specific sequences may be involved in copy number control as shown for the REP3 gene function in 2 μm circle replication (14) and in determining compartmentalization of the DNA before and after replication and at cell division. Some of these functions might be located in the RE domain and as they may be more cell and species specific than the basic function of replicating DNA the RE domains may show a greater divergence between different ARSs. This could explain the weaker ARS phenotype of human and Drosophila (unpublished data) ARSs in yeast. It will be interesting to replace a human RE domain with a yeast RE domain and examine the effect on the human ARS phenotype to test this idea. Our functional and sequence analyses of the RE domain in human ARSs have

indicated several conserved sequences. These sequences are clustered in the ARS1 RE domain, which is functionally defined by a short DNA segment, and dispersed in the ARS2 RE domain which is functionally defined by a relatively long DNA segment. These results support the notion that the blocks of homology shared by the two RE domains are biologically significant. At present so little is known about the spatial organisation and regulation of replication in eukaryotes that we can only speculate on the possible detailed biological functions of ARSs. It is possible that heterologous ARSs are only relevant for replication in yeast and occur fortuitously in these other eukaryotic genomes. This is less likely if the sequence requirements for ARSs activity are complex. This paper shows that two spatially distinct but associated regions of DNA are required for the ARS function of human DNA fragments. Furthermore two fragments chosen randomly from eight original isolates share homologous nucleotide sequences in these functional regions. The location of several conserved sequences with in a short stretch of DNA which correlate with functional activity is unlikely to have occurred by chance and suggests that the structural and sequence organisation of these regions has been conserved in the human genome. The isolation of ARSs from a mammalian genome will now allow the functional significance of ARSs isolated in yeast to be tested in the homologous system as tissue culture transformation systems are available. We are currently constructing appropriate vectors to assess replication efficiency of the sequences in human cells.

ACKNOWLEDGEMENTS

J.F.M. holds a fellowship from the Universidad Nacional Autonoma de Mexico. J.S.M is in reciept of an award from the M.R.C. for training in research methods, M.J.D. and M.F.T. are supported by Celltech Ltd. S.M.K. is an E.P.A. Cephalosporin Fund Research Fellow of the Royal Society of London. The work was funded by M.R.C. grant G8128390CB and Cancer Research Campaign grant Oxford BC3.

*To whom correspondence should be sent

REFERENCES

1. De Pamphlis, M. and Wasserman, P. (1980). Ann. Rev. Biochem. **49** 627-666.
2. Harland, R.M. (1981). TIBS, **60**, 71-74.
3. Blumenthal, A.B., Kriegstein, H.J. and Hogness, D.S. (1973). Cold Spring Harbour Symp. Quant. Biol., **38**, 205-223
4. Harland, R.M. and Laskey, R.A. (1980), Cell **21**, 761-771
5. Hamlin, J.L. (1978) Exptl. Cell Res. **112**, 225-234.

6. Taylor, J.H. Chromosoma (1977), **62**,291-303
7. Stinchcomb, D.T., Struhl, K. and Davis, R.W. (1979). Nature, 282, 39-43.
8. Hinnen, A. Hicks, J.B. and Fink, G.R. (1978) Proc. Nat. Acad. Sci.**75**, 1929-1932
9. Kingsman, A.J., Clarke, L., Mortimer, R.K. and Carbon, J.C. (1979) Gene, 7,141-152.
10. Struhl, K., Stinchcomb, D.T., Scherer, S. and Davis R.W. (1979) Proc. Nat. Acad. Sci.USA , **76**, 1035-1039.
11. Beach, D. Piper, M. and Shall, S. (1980) Nature **284**, 185-187.
12. Chan, C.S.M. and Tye, B. (1980) Proc. Nat. Acad. Sci. USA **77**, 6329-6333.
13. Stinchcomb, D.T., Thomas, M., Kelly, J.,Selker, E. and Davis, R.W. (1980) Proc. Nat. Acad. Sci.USA **77** 4559-4563.
14. Broach J.R., Li, Y-Y, Feldman, J., Jayaram,M., Abraham, J., Naysmith, K.A. and Hicks,J.B. (1982) Cold Spring Harbour Symp. Quant. Biol. **47**, 1165-1175.
15. Kearsey, S. (1983) EMBO J. **2**, 1571-1577
16. Kiss, G.B., Amin, A.A.and Perlman, R.E.(1981) Mol. Cell Biol. 1, 535-543.
17. Zakian, V.A. (1981) Proc. Nat. Acad. Sci.USA **78**,3128-3133
18. Petes, T.D. and Williamson, D.H. (1975) Exptl. Cell Res. **95** 103-110.
19. Celniker, S.E. and Campbell, J.L. (1982) Cell, **31**, 201-213.
20. Jazwinski, S.M., Niedswieda, A. and Edelman, G.M. (1983) J.Biol. Chem. **258**, 2754-2757.
21. Fangman, W.L., Hice, R.H. and Chlebowicz-Sledziewska, E. (1983) Cell, **32**, 831-838.
22. Tschumper, G. and Carbon, J. (1980) Gene **10**, 157-166.
23. Stinchcomb, D.T., Mann, C., Selker, E. and Davis, R.W. (1981) In, ICN-UCLA Symposium on Molecular and Cellular Biology. eds. D.S. Ray and C.F. Fox, Academic Press vol. 22, pp.473-488.
24. Hartley, J.L. and Donelson, J.E. (1980) Nature **286**, 860-865.
25. Tschumper, G. and Carbon, J. (1982) J. Mol. Biol. **156**,293-303.
26. Tschumper, G. and Carbon, J. (1981) In, ICN-UCLA Symposium on Molecular and Cellular Biology, eds. D.S. Ray and C.F. Fox, Academic Press vol. 22, pp.489-500.
27. Feldman, H., Olah, J. and Friedenreich,H. (1981) Nucl. Acid Res. **9**, 2949-2959
28. Miller, J.H. (1972) Experiments in Molecular genetics, Cold Spring Harbour Laboratory, Cold Spring Harbour New York p 433.
29. Hawthorne, D.C. and Mortimer, R.K. (1960) Genetics **45**, 1085-1110.
30. Chinault, A.C. and Carbon, J.A. (1979) Gene 5, 111-126.
31. Holmes, D.S. and Quigley, F.A. (1981) Anal. Biochem. **114**, 193-197.
32. Cryer, D.R., Eccleshall, R. and Marmur, J. (1975) in Prescott, D.M. (ed.), Methods in Cell Biology, Academic Press, N.Y. vol.12, 39-44.
33. Fergusson, J., Groppe, J.C. and Reed S.I. (1981), Gene **16**, 191-197.
34. Cook, P.R. and Brazell, I.A. (1980) Nucl. Acids Res. **8**, 2895-2966.
35. Southern, E.M. (1975) J.Mol. Biol.**98**, 503-517.
36. Weinstock, R., Sweet, R., Weiss, M., Cedar, H. and Axel R. (1978) Proc. Natl. Acad. Sci. USA **75**,1299-1303.
37. Dobson, M.J., Tuite, M.F., Roberts, N.A., Kingsman, A.J., Kingsman, S.M., Perkins, R.E., Conroy, S.C., Dunbar, B. and Fothergill, L.A. (1982) Nucl. Acid Res. **10**, 2625-2657.
38. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Nat. Acad. Sci.USA. **74**, 5463-5467.
39. Messing, J. and Viera,J. (1982) Gene **19**,269-276.

40. Struhl, K. and Davis, R.W. (1980) J. Mol. Biol.136, 309-332.
41. Bolivar, F. (1978) Gene 4, 121-136
42. Drouin, J. (1980) J. Mol. Biol. 140, 15-27.
43. Tikhomirova, L.P., Kryuokov, V.M.< Strizhov, N.I. and Bayev, A.A. (1983) Molec. Gen. Genet. 189, 479-484.