Determination of the Functional Domain of a Mouse Autonomous Replicating Sequence

Chiharu HAYASHI,[a] Hiromichi FUJINO,[a] Masanori OGATA,[c] Yoshinori SATO,[c] Sanae M. M. IGUCHI-ARIGA,[b] and Hiroyoshi ARIGA*,[a]

Faculty of Pharmaceutical Sciences, Hokkaido University,[a] Kita 12, Nishi 6, Kita-ku, Sapporo 060, Japan, College of Medical Technology, Hokkaido University,[b] Kita 12, Nishi 5, Kita-ku, Sapporo 060, Japan and Research Institute, Daiichi Seiyaku Co., Ltd.,[c] 1-16-13 Kita-kasai, Edogawa-ku, Tokyo 134, Japan.
Received December 27, 1996; accepted March 1, 1997

We previously isolated from mouse cells an autonomous replicating sequence (ARS) ARS65 (Ariga, Itani and Iguchi-Ariga, Mol. Cell. Biol. 7, 1—6, 1987). Here we report the nucleotide sequence of ARS65. The sequence from BglII to EcoRI sites cloned as ARS was 2658 bp long. There exist three interesting domains: a TA repeat, a *myc* like box (essential sequence for c-*myc* ARS), and a T rich region. Cloned DNAs containing various segments of pARS65 were transfected to rat 3Y1 cells together with the hygromycinB resistance expression vector, and hygromycinB resistant clones were isolated. Established cell lines transfected with plasmids carrying either a *myc*-like box or a T rich region harbored the replicated plasmids, indicating that these two elements are necessary for the ARS function of pARS65.

Key words ARS; c-*myc*; replication

There are generally believed to be about 10^5—10^6 initiation regions of DNA replication (*ori*) in mammalian cells,[1] and several mapping protocols including BrdU/PCR and 2D gel methods have been established to identify them.[2,3] This method identified the several *oris* in mammalian genomic DNAs that include the genes of DHFR,[4,5] c-*myc*,[6] immunoglobulin heavy chain (IgH),[7,8] hsp70,[9] lamin B,[10] β-globin,[11,12] and adenosine deaminase.[13] Another approach has been to isolate autonomous replicating sequences (ARS), which are *ori* candidates. Indeed, in *S. cerevisiae* some of the ARSs identified functioned as *oris* in chromosomes.[14] In mammalian cells, several ARS including those from the genes of c-*myc*,[15,16] IgH,[9] and hsp70,[7] also, worked as an *ori in vivo*. From the characterization of ARSs and *oris*, the *ori* consensus structure seems to be a structure that aligns the presence of transcriptional recognition sequences, an AT rich sequence, and bent DNA.

We previously cloned an ARS, pARS65, from mouse cells which was able to replicate episomally in several kinds of mammalian cells.[17] In this manuscript, we show the nucleotide sequence of pARS65, and identify the domain necessary for the maintenance and replication of the plasmid in the cells.

MATERIALS AND METHODS

Cells Rat 3Y1 cells were cultured in Dulbecco-modified Eagle medium supplemented with 10% fetal calf serum.

Establishment of Cell Lines Resistant to Hygromycin B Rat 3Y1 cells (1×10^7 in a 10 cm dish) were transfected with 19 μg of test plasmids together with 1 μg of a hygromycinB resistance-expression vector (pLTRHyg) by calcium phosphate precipitation method.[18] Five hours after transfection, the cells were boosted with 25% glycerol and then cultured for 2 d. The cells were replated at the density of 1×10^6 cells per 10 cm dish and cultured in the medium containing 20 μg/ml hygromycinB. The medium was changed every 4—5 d, and colonies of hygromycin B-resistant cells were isolated in about 2 weeks.

Analyses of Cell Lines Carrying Autonomously Replicating Plasmids Low molecular weight DNAs were isolated from the established cells by the method of Hirt,[19] digested with DpnI or MboI, blotted on a nitrocellulose filter by the method of Southern,[20] and hybridized with a ^{32}P-labelled pUC19 probe as described.[17]

The nucleotide sequence data in this paper have been submitted to the EMBL Data library under accession number X70989.

RESULTS AND DISCUSSION

Nucleotide Sequence and Structure of pARS65 The fragment from BglII to EcoRI sites in pARS65[17] was subcloned in pUC19, and the whole nucleotide sequence was determined by the dideoxy chain termination method (Fig. 1). The sequence is 2658 bp long, and contains three interesting regions (see Fig. 2): A 52 bp TA repeat (nucleotide 1564 to 1615); a T rich region (1991 to 2069); two *myc*-like boxes, ACTCTTATA (1756—1764) and ACTCTTATA (1787—1795). The two *myc*-like boxes are homologous to TCTCTTATA, the core of the 21 bp long element essential to the ARS function that we identified in the HindIII-PstI fragment (H-P) upstream from the c-*myc* gene.[15,16] The (H-P) region in the c-*myc* gene also possesses a TA-rich sequence beside TCTCTTATA . The repartition of these sequences is similar between the two ARSs. Therefore, we tried to determine the functional domains of pARS65 using a long term replication assay.

Determination of the Functional Domain for Plasmid Replication in Rat 3Y1 Cells Portions of pARS65 were subcloned into pUC19 and transfected to rat 3Y1 cells with the hygromycinB resistance (Hyg^r) expression vector, pLTRHyg[22] by the calcium phosphate precipitation method1.[18] The entire sequence of pARS65 was also recloned in pUC19 and called pUC-ARS65. Test plasmids used for transfection are shown in Fig. 2. Nineteen micrograms of test plasmid DNA and one microgram pLTRHyg

* To whom correspondence should be addressed.

```
1          11         21         31         41         51         61         71         81
AGATCTATTTAAGCAAAATCTGAGAAATAGTATGTTCCTCTTGTGTCTCTGCTGATATAATTTCAGTGTGTATAATAAGATTATTAAAAA
91
ATGACAGTGGACACATAGGTGCACACTATAGATTAAAAATCAATACAACTTTGAAAGAAAGAAAGAAGTAAGCTAGTTGAGGGAAATTAC
181
AAAACTGAAAATTCATTAATTCCAAATTTTAAGGCAGACTGCTTCCCAGAAAGCATACCACTGAGCAATCGCTTTTTGAATATGTGATTT
271
CTATAAAGTACATTGTCTTTTGCTAGCGCATATTCTGCCTGTTGACACTAGGACAGTATTTCACTGGGAGCAAATAAGTGCTATTTCTTT
361
AAAATACACTTTACTTCAAAAAATGGGTGTCTTGGCCACATTTCCATGCAGTGTCTTTAACACATTATATATCACTTTGACCTTCATCTT
451
GACTCATTATTGTACAGCAAAAGAAAGCATAATTTGTTGTTGTATGGCTTATGATGTGATGATAATAGAGCTTATTTTTACACAGAATTT
541
TATGTAACATTGTAACCGCAAGGCAATTAATAAATAAAAAGTTGTTGGGCATATATATCAACAAACCCTAATTACATTTGATACTTCATT
631
AGTATCATTTTCTATGGTAGTTTCTTTGATTTGGCTCACATAGTCTAATTAGGTCCTAATACAGTCTTCAACTGTGCACCTGGCTCTAAC
721
CATCCATGCAATTTTGTGAACTATACACCAGACTCTAACCACCTATGCACTGTTGTAAGCTGTGTGCCAGGCTCTAAACATCCATCTACT
811
ATAGGGAACTGTGCAACAGGTCTGACCTACCCGTACACTATTGTGAACTTTGCATTGGTGCCAATCAGAATCTTGCCTATCAATGTGGGA
901
GTCGGCTTTTCAGTATTACAAGAAAATAGCTGAGTTGTAGAGCATTATAATTTTTTCGAATTAATTTTGCTCATATTTTGAATTAATGTG
991
ACAGACGTTCTGTGTGGCCTAGTTCATGCTTGAGTCACACCATGACCAGTAGCAATGGTTGTAATCTTAGGAGGTGATTGGCTATAACTA
1081
TTAGGAGCAGGGAGAAAGAATGTAGCCATACCCTATTATACTTTTATTAAAAATAACATTCTTATGAAAACTACCAGTGGATTAAATGTG
1171
AATTACTAAGGGACTGAGAGAGACAGATTTCATATCAACAACTGCTGGATGACAGCCTATATAGTTCTACTTTGTAAAGACTCACTTCTC
1261
TACCAGTACTACCTTGAGTCCAAGTATACCTTGGGGATCACAAACCACATTCAAAAACAAAGGAAAACTCTTTGTTTTAGATTGAGGAAA
1351
AAAAATTTCCTCCAAATGTATTCTGTGTAAAGCCTTTACAATGTTTTTTAATTAAGCCTTTACAATTAAAAACAAAGTCCAGAGTCTGTG
1441
ACTTTTAGTCATCTGACGTTCAACAAGAACACATGCAAAACAGAGTACTGATTAGAAGACTCATATCATCAAGGTGTAAAGCAGTTGTAC
1531                              1564                                                    1615
ATTTGTGCTTTTTGAAGATAGTCTACTGGTAAATATATATATATATATATATATATATATATATATATATATATATATATATATAATTGT
1621                                                  TA repeats
ATCAGAAATAGGAATAAAGTAACAAATGTAACAGCTAGCCAAGGTTCTGAAGAGTAGCCCCAACTAGGAAGTAGAGTGGGAAAGGACAGA
1711                                                  1756     1764                    1787     1795
TGACAACAGACTGTGAGGGTAACTGGTATTTAATTGCTAATGATAACTCTTATATTGGTGTTCAATAATTAATGTTACTCTTATACTTGC
1801                                                  myc like box                     myc like box
TCTATACAAGCAACCGACATGACAGCATTCTATTAATTATTTAGAAGCATGAACCATATAGCATATTAATATCAGACTCAAGTTGTTGAG
1891
AGAGAGAGTTGCTTCTGAATGTTCACAGACAATAATGCAGCAACTAACAAACATCTTTGTTTCTGAGAGAAAAAATAGTACCTTATCAAG
1981       1992
AGAGTTAAGCCATTTTGTTTTGTTTTGAATTTTTTGTTTAATTTTTTTTTGGGGGGGGTTGTTTTGTTTTGTTTTCCGGGAAAGGGTTTC
2071   2078                          T rich region
TCTGTATAGCCATGGCTGTCCTGGAACTCACTCTGTAGACCAGGCTGGCCTCGAACTCAGAAATCCTCCTGCCTCTGCCTCCCAAGTGCT
2161
GGGATTAAAGGCATGCGCCATCACTGCCTGGAGTAAACCATTTTCATAATTGAAGAAATAGAACGATTGCTTCTGAGCTCTCTGAGTCCT
2251
TTAATCCATGGTTAATTTACATAACAATGCTTTATAGATGCCATATACTTTCCAGTGCCATCTAAATATCTTTTTTTTTTCCTGTCTCTC
2341
AATGCAGGTTTGTGATGTATTAAGATATATAAGATGCATTCATCTTTTTGAAATCTGTCCAAAAAATCAGTTTCTCTAAGTATGATAGGA
2431
TTATCCACTAAAACACCAGCCATGTTCAGTCAGTCATGTGGCTTATAGGGAGGAACGTGGCCCTGATGATACATGGCATTTTAGCAGTTC
2521
TCCTGATCCAAAGTAAAATGATTGATCCCCAATGAAGCCCACCACAGTGTTTTACATGACTGGAACAACATTTCATGATATCGAGATTTA
2611
ACAGGGTTACAATTTTCTCTGATTAGGAAATTCTTGAGACTAGAATTC
```

Fig. 1. Nucleotide Sequence of the Mouse Autonomous Replicating Sequence, pARS65

TA repeat, *myc*-like box, and T rich region are underlined.

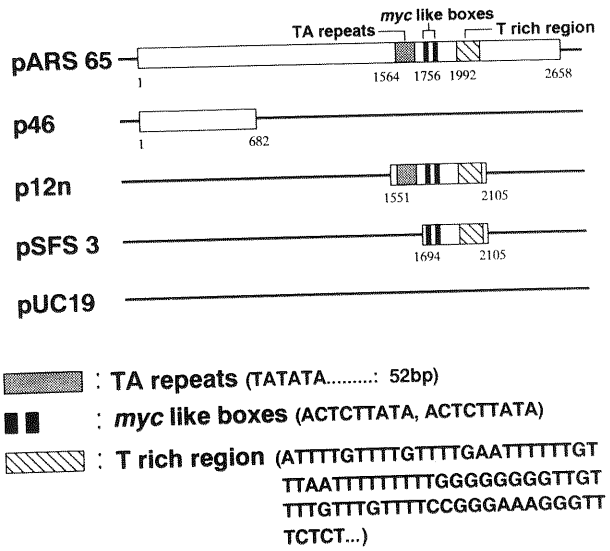

Fig. 2. Schematic Representation of the Structure of pARS65 and of the Plasmids Used in the Transfections to Establish Cell Lines

Table 1. Establishment of Cell Lines Harboring the Plasmid DNAs

Test plasmid	Number of cell lines			Efficiency (%)
	Established	DNA examined	Containing plasmids	
pARS65	10	7	3	42
pUC-ARS65	9	6	2	33
p46	16	15	0	0
p12n	24	20	1	5
pSFS3	24	18	2	11
pUC19	13	11	0	0

Cell lines resistant to hygromycinB were established after co-transfection of various DNAs and a hygromycinB resistance expression vector to rat 3Y1 cells. Five micrograms of total DNAs were then digested with various restriction enzymes and subjected to Southern blotting. pUCARS65 is a subclone of pARS65, in which the BglII–EcoR I mouse DNA fragment of pARS65 was inserted into the BamH I–EcoR I sites of pUC19.

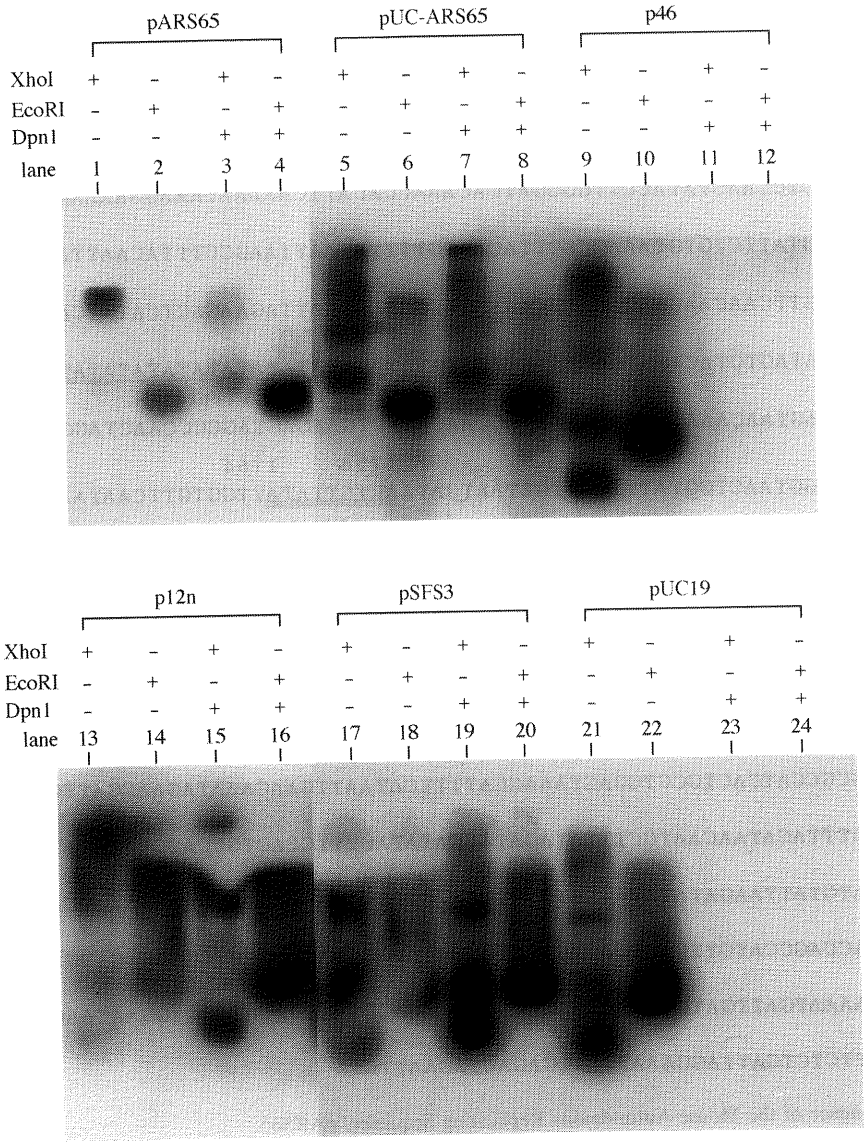

Fig. 3. Southern Blot Analysis of Plasmids Replicated in the Established Cell Lines

Total DNAs were extracted from the established cell lines, digested with EcoRI or XhoI with or without DpnI, and subjected to Southern blot analysis. Hybridization was carried out with a ^{32}P-labeled pUC19 probe.

were transfected to 50—60% confluent rat 3Y1 cells in 10 cm dishes. Two days after transfection, the transfected cells were cultured in medium containing hygromycinB (75 μg/ml) and cell lines resistant to hygromycinB were established (Table 1). Total DNAs were extracted from the cells. Five micrograms thereof were left untreated or digested with appropriate restriction enzymes, transferred to nitrocellulose filter as described by Southern,[20] and hybridized with pUC19-specific probe under stringent conditions.[17] To distinguish the plasmid from the integrated form of the transfected DNAs, restriction enzymes were used, which give one cut to each plasmid DNA (EcoRI), or which release the inserted fragments, or which do not digest the plasmid DNAs (XhoI). While p46 and pUC19 were not harbored, plasmid DNAs were detected in the cells transfected with pARS65, pUC-ARS65, p12n and pSFS3 (Fig. 3, Table 1). These plasmid DNAs were resistant to DpnI and sensitive to MboI digestions (data not shown), indicating that the plasmid DNAs indeed replicated in the established cell lines. Efficiency of the cells harboring the plasmid DNA is not as high as that in our previous report, in which mouse FM3Atk$^-$ cells in suspension culture were transfected with pARS65-tk by the liposome method[17,23] and cultured in HAT medium. A possible reason for this discrepancy is that pARS65-tk in FM3Atk$^-$ had a selective advantage for maintenance, while the plasmid found to replicate in this report had no relation with resistance to hygromycinB.

It is interesting that 52 base pairs TA repeats are not necessary for the plasmid to replicate and be maintained in the cells. In contrast, ARS function of pARS65 seems to reside in a region comprising the *myc*-like boxes and T rich region. Although it is not clear from this experiment which of these two is essential for ARS function, it is possible from previous results concerning the ARS from the c-*myc* gene[15,16,21] that *myc*-like boxes are essential for ARS, but that both *myc*-like boxes and T rich region are necessary for the maintenance of plasmid over the long term. Although the efficiency to establish cell lines harboring plasmid DNA in an episomal state varied among plasmids tested, copy numbers of each plasmid replicated in the established cell lines were at almost the same level, which suggests that the region other than *myc*-like boxes and the T rich region affects the stability of plasmid. The ARS sequence present in c-*myc* gene gave rise to much higher copy number of plasmids replicated in the cells than that from ARS65.[15,17] The essential sequence of c-*myc* ARS, the 21 base pairs, has an AT rich sequence just beside 21 base pair, while the T rich region is separated to *myc* like boxes 240 base pairs in length in ARS65. This distance might give less active replication efficiency of pARS65. These kinds of cell lines provide good model systems to identify the molecular mechanism of DNA replication in mammalian cells.

Acknowledgment We gratefully acknowledge Ivo Galli for critical reading of this manuscript. This work was supported by the grants from the Ministry of the Education, Science, Sports and Culture in Japan and the Suhara Memorial Foundation.

REFERENCES

1) Huberman J. A., Riggs A. D., *J. Mol. Biol.*, **32**, 327—341 (1968).
2) Vassilev L., Johnson E. M., *Nucleic Acids Res.*, **17**, 7693—7705 (1989).
3) Nawotka K. A., Huberman J. A., *Mol. Cell. Biol.*, **8**, 843—853 (1988).
4) Dijkwel P. A., Hamlin J. L., *Mol. Cell. Biol.*, **12**, 3715—3722, (1922).
5) Vassilev L. T., Burhans W. C., DePamphilis M. L., *Mol. Cell. Biol.*, **10**, 4685—4689 (1990).
6) Vassilev L., Johnson E. M., *Mol. Cell. Biol.*, **10**, 4899—4904 (1990).
7) Iguchi-Ariga S. M. M., Ogawa N., Ariga H., *Biochim. Biophys. Acta*, **1172**, 73—81 (1993).
8) Ariizumi K., Wang Z., Tucker P. W., *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 3695—3699 (1993).
9) Taira T., Iguchi-Ariga S. M. M., Ariga H., *Mol. Cell. Biol.*, **14**, 6386—6397 (1994).
10) Biamonti G., Perini G., Weighardt F., Riva S., Giacca M., Norio P., Zentilin L., Diviacco S., Dimitrova D., Falaschi A., *Chromosoma*, **102**, S24—31 (1992).
11) Kitsberg D., Selig S., Keshet I., Ceder H., *Nature*, **366**, 506—507 (1993).
12) Aladjem M. I., Groudine M., Brody L. L., Dieken E. S., Fournier R. E., Wahl G. M., *Science*, **270**, 815—819 (1995).
13) Carroll S. M., DeRose M. L., Kolman J. I., Nonet G. H., Kelly R. E., Wahl G. M., *Mol. Cell. Biol.*, **13**, 2971—2981 (1993).
14) Umek R. M., Linskens M. H. K., Kowalski D., Huberman J. A., *Biochim. Biophys. Acta*, **1007**, 1—14 (1989).
15) Iguchi-Ariga S. M. M., Okazaki T., Itani T., Ogata M., Sato Y., Ariga H., *EMBO J.*, **7**, 3135—3142 (1988).
16) Sudo K., Ogata M., Sato Y., Iguchi-Ariga S. M. M., Ariga H., *Nucleic Acids Res.*, **18**, 5425—5432 (1990).
17) Ariga H., Itani T., Iguchi-Ariga S. M. M., *Mol. Cell. Biol.*, **7**, 1—8 (1987).
18) Graham F. L., van der Eb A. J., *Virology*, **52**, 456—467 (1973).
19) Hirt B., *J. Mol. Biol.*, **26**, 365—369 (1967).
20) Southern E.M., *J. Mol. Biol.*, **93**, 503—517 (1975).
21) Ariga H., Imamura Y., Iguchi-Ariga S. M. M., *EMBO J.*, **8**, 4273—4279 (1989).
22) Blochlinger K., Diggelmann H., *Mol. Cell. Biol.*, **4**, 2929—2931 (1984).
23) Itani T., Ariga H., Yamaguchi N., Tadakuma T., Yasuda T., *Gene*, **56**, 267—276 (1987).