Nucleic Acids Research, Vol. 20, No. 13 3333–3339

# The AT-rich tract of the SV40 *ori* core: negative synergism and specific recognition by single stranded and duplex DNA binding proteins

Ivo Galli, Sanae M.M.Iguchi-Ariga[1] and Hiroyoshi Ariga*
Faculty of Pharmaceutical Sciences and [1]College of Medical Technology, Hokkaido University, Kita-ku, N12 W6, Sapporo 060, Japan

Received April 16, 1992; Revised and Accepted June 15, 1992

## ABSTRACT

**The SV40 origin of replication comprises a run of thymine and adenine residues. Integrity of this AT-rich sequence is known to be essential for replication. We set out to study whether or not these elements can work synergistically to sustain replication. Quite surprisingly, additional copies of the AT stretch linked to a functional SV40 *ori* core dramatically reduce its replication in CosI cells, probably by creating some physical block. Interestingly, the same inhibiting effect can be observed with the addition in *cis* of the yeast ARS consensus, which is homologous to the SV40 AT stretch. This modulation is possibly due to the action of cellular factors that recognize either of the two sequences. In fact, we demonstrate the existence of factor(s) in CosI crude nuclear extracts that in *vitro* can specifically bind to either of them. Moreover, we show that these sequence-specific factor(s) (MW about 50 kDa), named SOAP, recognize both single (T-rich strand) and double stranded forms of the AT tracts. Binding to single stranded AT stretches can be specifically inhibited by the corresponding duplex form, but not *vice versa*.**

## INTRODUCTION

The SV40 origin of replication (*ori*) has been mapped as a 63 bp long core sequence (1–5). It is flanked by two auxiliary sequences that augment replication, but are not indispensable (4, 6–8). Among other things, they facilitate DNA unwinding (9). On its early-genes side, there is a palindromic region comprising the second binding site for SV40 T antigen (Tag), followed by a short inverted repeat sequence (an imperfect, minor palindrome) and the actual origin of bidirectional replication (10–12). The late-genes side of the *ori* core consists of a 15 bp long AT-rich tract that also provides a TATA box for the early promoter (12, 13). Point mutations and deletions as well as insertions within the AT stretch were reported to reduce replication (3, 14).

AT stretches are a characteristic of many viral, as well as of several prokaryotic and eukaryotic origins of replication (15–18). In particular, papovaviruses present an uninterrupted tract of at least six thymine residues. Nevertheless, the function of these AT tracts is not totally clear. On one side, they are likely to affect the structure of the DNA by their weak base pairing and by their bending potential (14). The two properties are most probably a very basic mechanism in regulation of replication. In fact, the SV40 *ori* AT stretch was found to contain two of the major primer start sites for replication (10). On the other side, the AT stretches are target sequences for cellular factors (19, 20). With this regard, the accurate study of viral AT stretches may provide some useful insight in the mechanisms that underlie cellular replication.

One interesting question to be addressed is whether or not multimerized AT stretches can work synergistically in implementing SV40 *ori* core replication. In fact, it is known that in several cases multimerization of binding sites brings about stronger transcriptional activity (21, and references therein). Would similar considerations apply to replication? The new finding, shown in this work, is that the number of AT stretches present beside a functional SV40 *ori* core affects the extent of replication in a negative way. Moreover, the yeast ARS consensus, reminiscent of the SV40 AT stretch, is also able to inhibit replication of the SV40 *ori* core in a very similar way. Modulation is probably due to the action of cellular factor(s) that recognize either sequence. In vitro, we demonstrate the existence of proteins in crude CosI nuclear extracts that indeed bind to either the SV40 AT stretch or to the yeast ARS consensus. In addition, we observed that these factor(s) (about 50 kDa) bind to both single (T-rich) and double stranded forms of the AT tracts.

## MATERIALS AND METHODS

### Cell culture and transfection

CosI cells were cultured in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal calf serum (Boehringer Mannheim), at 37°C under 5% $CO_2$. For *in vivo* replication assays, 50% confluent cells (in 6 cm dishes) were transfected by the calcium phosphate technique (22). The amounts of DNA were 750 ng for the test plasmids and 100 ng for the internal

* To whom correspondence should be addressed

control CibiH (see below). Four to five hours after transfection, the cells were shocked with 20% glycerol in medium for 5 minutes at room temperature, then incubated for 48 hours before harvesting. Low molecular weight DNAs were recovered by the Hirt procedure (23), digested with DpnI enzyme to eliminate non-replicated, methylated DNA, linearized with an appropriate restriction enzyme (HindIII) and analyzed by Southern assays (24). The probe was vector-specific, in this case pUC19 specific, and was obtained by labelling with $^{32}$P-dCTP nascent, random primed fragments. In all replication assays an internal control was included in order to standardize the efficiency of transfection. The internal control consists of a pUC-derived plasmid (CibiH) that is shorter than the test plasmids and can be therefore easily separated electrophoretically (in the Southern assays) upon linearization with HindIII. CibiH was grown in the dam$^-$ *E.Coli* strain GM113, so that it cannot be degraded by the DpnI enzyme.

### Plasmids

All plasmids were prepared either by inserting synthetic oligonucleotides into the corresponding restriction sites of pUC19, or by DNA manipulations that were performed according to standard techniques (25).

### Oligonucleotides

Oligonucleotides were synthesized chemically with an Applied Biosystems DNA synthesizer, cleaved from their support by usual procedures and purified on a HPLC cartridge (Applied Biosystems). Their sequences are the following (from 5′ to 3′):

AT+ GATCCATTTTTTTTATTTAT
AT− CTAGATAAATAAAAAAAATG
GC+ CTAGAGGCCGAGGCCGCCTCGGCCTCTGAGC-TATTCCAGAAGTAGTGAG
GC− TCGACTCACTACTTCTGGAATAGCTCAG-AGGCCGAGGCGGCCTCGGCCT
Y+ GATCCATTTTATGTTTATCTAGAG
Y− GATCCTCTAGATAAACATAAAATG
PEB+ GATCCTGCGGTTGAATAGTCACCTCTG
PEB− GATCCAGAGGTGACTATTCAACCGCAG

For the band shift probes, annealed double stranded oligonucleotides were end-labelled with [$\alpha$-$^{32}$P]dCTP by Klenow enzyme fill-in reaction. Single stranded probes were prepared by T4 polynucleotide kinase phosphorylation with [$\gamma$-$^{32}$P]ATP.

### Nuclear extracts

Nuclear extracts were prepared from cells in exponential growth as described by Dignam and his colleagues (26), with minor modifications (27).

### Band shift assays

Binding of about 8 $\mu$g CosI crude nuclear extract to 10000 cpm probe (aprox. 0.35 pmoles) was performed in 15 $\mu$l with 1 mM EDTA, 4% Ficoll-400, 1 mM DTT, 4 mg/ml Bovine Serum Albumin, 50 mM KCl, 2 $\mu$g poly(dI:dC), on ice, for 20 minutes. The mixtures were then separated at 4°C through a 4% polyacrylamide gel (19:1 crosslink, 0.25×TBE Buffer) (28, 29).

### Ultraviolet light-mediated crosslinking assays

Eight $\mu$g CosI crude nuclear extract were allowed to bind to probes (60000 cpm) under the same conditions as for band shifts (see above). Following a known protocol (30), the proteins interacting with the probes were then covalently linked to the DNAs by irradiation with UV light (254 nm) on ice for 30 min. After addition of an equal volume of 2× Laemmli Buffer and denaturation at 70°C for 5 min, the mixtures were separated through a 10% polyacrylamide (29:1 crosslink) gel containing 0.1% SDS (31).

### Southwestern assays

About 20−40 $\mu$g nuclear extract were separated through a 10% SDS polyacrylamide gel (31), blotted to a nitrocellulose filter in 20 mM Tris-150 mM glycine for 16 hr at 17 V, 4°C. The filter was then blocked with 10 mM Tris (pH 7.4)-1 mM DTT-5% non-fat milk powder (Snow Brand) at 4°C for 1 hr. The binding reaction of 1 hr was performed at 4°C in a buffer containing 10 mM Hepes (pH 7.9), 50 mM NaCl, 10 mM $MgCl_2$, 0.1 mM EDTA, 1 mM DTT, 10 mg/ml poly(dI:dC), 0.25% non-fat milk powder, with $10^6$ cpm/ml probe. The filter was washed three times at 4°C for a total of two hours with an identical buffer except for the concentration of NaCl (0.2 M) (32).

## RESULTS

The SV40 origin of replication (*ori*) core consists of two fundamental units. On its late-genes side, there is an adenine and thymine-rich sequence (AT stretch) that appears to be a structure common to most replication origins of the papovaviruses (3) as well as of several cellular replication systems (Fig.1a). On its early-genes side, there is a relatively GC-rich sequence that comprises a palindromic binding site for SV40 T antigen and a short imperfect palindrome (Tag/IR unit, Fig 1b) (12).

In order to analyze possible synergistic effects of the AT stretches, we synthesized two pairs of oligonucleotides corresponding to these two fundamental units (see 'Materials and Methods'). A novelty lies in the fact that we introduced a useful XbaI restriction site at their junction by simply changing two nucleotides in the sequence (Fig. 1b), without altering the relative

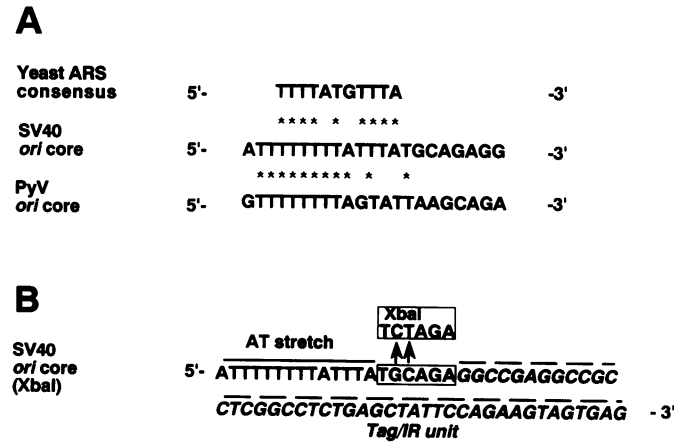


**Fig.1.** Sequence comparisons and definitions. A, the SV40 *ori* core adenine and thymine-rich sequence (AT stretch), and similar sequences. B, full sequence of the SV40 *ori* core. The AT stretch is marked with a continuous line. Tag/IR unit (italics, dashed line) indicates the binding site for T antigen and the short inverted repeat (imperfect palindrome) on the early genes side. In this work, we used a mutant sequence with two point mutations (GC to CT) at the junction between the AT stretch and the Tag/IR unit. The mutation yields a novel XbaI restriction site.

distance. First, we verified the replicating activity of this new system. Several plasmid DNAs representing the AT stretch and the Tag/IR unit linked in different orientations and positions were prepared using these oligonucleotides, and transfected to CosI cells together with a DpnI resistant control plasmid (CibiH) for evaluation of the efficiency of transfection. CosI cells constitutively produce SV40 T antigen and can therefore support replication of wild type core origins. Low molecular weight DNA was extracted by the Hirt procedure (23), digested with DpnI enzyme to degrade non-replicated, methylated DNA, linearized with HindIII and analyzed by a Southern blotting assay (24) with a vector-specific probe. As shown in Fig. 2, the reconstituted

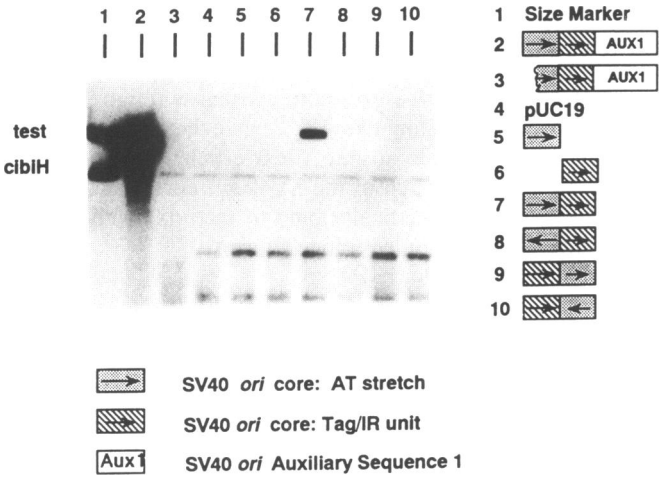


**Fig. 2**. Effects of inversion and misplacements of the AT stretch from the rest of the *ori* core. Shaded boxes indicate the SV40 *ori* core AT stretch; boxes with dense, slant lines indicate the Tag/IR unit; white boxes correspond to Auxiliary Sequence I (early genes side). The relative orientations are shown by arrows. The reconstituted wild type SV40 *ori* core (XbaI) corresponds to lane 7. CibiH is the internal control for transfection efficiency (see 'Materials and Methods'). Size markers (lane 1): about 1 ng pUC19 and CibiH, respectively.

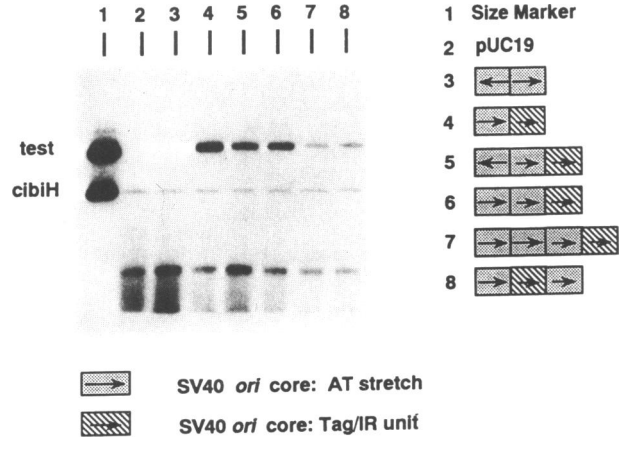


**Fig. 3**. Effects of multimerization of AT stretches. Shaded boxes indicate the SV40 *ori* core AT stretch, while boxes with dense, slant lines indicate the Tag/IR unit. The relative orientations are shown by arrows. The reconstituted wild type SV40 *ori* core (+XbaI) corresponds to lane 4. CibiH is the internal control for transfection efficiency (see 'Materials and Methods'). Size markers (lane 1): about 1 ng pUC19 and CibiH, respectively.

SV40 *ori* core (with the new XbaI site) can replicate (lane 7), albeit less efficiently than a longer origin region that also contains Auxiliary Sequence I (lane 2) (4, 6, 7). By contrast, the mutants bearing the AT stretch in the opposite orientation or downstream from the Tag/IR unit failed to replicate altogether (lanes 8–10). We can therefore confirm that the introduction of a new, practical restriction site into the SV40 *ori* core has little effect on replication. Besides, we can validate the findings that the juxtaposition of an AT stretch and a Tag/IR unit alone is not enough for granting replication, and that an appropriate positioning of the units that make up the *ori* core is necessary for function (3, 14).

## Effect of multimerization of the AT stretch

Next, we investigated the effect of joining more than one AT stretch to a Tag/IR unit. As shown in Fig. 3, the more AT stretches are present beside an *ori* core, the less efficient becomes replication. Interestingly, it is apparent that when an additional AT stretch is located far away (downstream from the Tag/IR unit; Fig. 3, lane 7) replication is even weaker than when it is located close to the functional AT stretch (that is, the one belonging to the *ori* core). In this case, the orientation of the additional AT stretch does not matter (lanes 4 and 5). The hypothesis arises that the presence of additional AT stretches reduces replication by creating a physical block that interferes with the normal course of replication. The nature and mechanism of this negative synergism remain obscure so far.

## SV40 ori core replication is affected by the yeast ARS consensus, which corresponds to a mutated SV40 AT stretch

As shown in Fig 1a, the yeast consensus sequence for a minimal origin of replication (33–35) closely resembles to the SV40 *ori* AT stretch. We examined its effects on SV40 replication by substituting it for the SV40 AT stretch, as well as by adding one or three copies upstream from a wild type SV40 *ori* core. The yeast sequence (see 'Materials and Methods' for a description

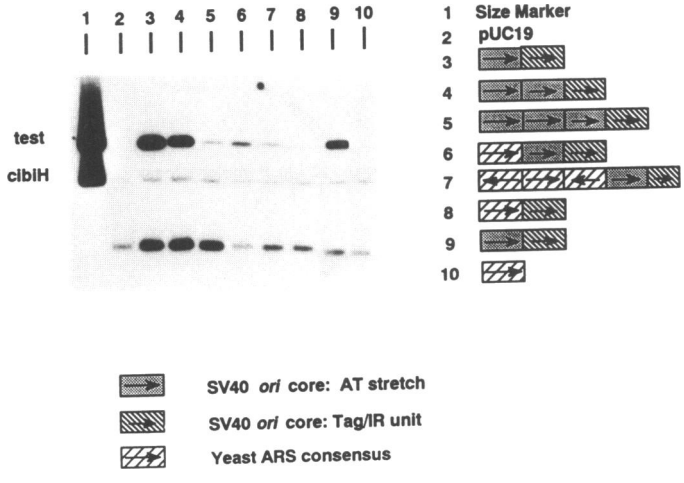


**Fig. 4**. Substitution of a yeast ARS consensus sequence (ATTTTATGTTTA-TCTAGA) for the SV40 *ori* core AT stretch (ATTTTTTTTATTTATCTAGA). Shaded boxes indicate the SV40 AT stretch; boxes with dense, slant lines (left to right) indicate the Tag/IR unit; boxes with sparse, slant lines (right to left) indicate the yeast ARS consensus, which in lane 8 is linked to the Tag/IR unit in the same way as the SV40 AT stretch (lanes 3 and 9), i.e. with an XbaI linker. Relative orientations are indicated by arrows. CibiH is the internal control for transfection efficiency (see 'Materials and Methods').

of the oligonucleotides) was linked to the SV40 Tag/IR unit at a distance corresponding exactly with that from the SV40 AT stretch. The data displayed in Fig.4 indicate that while the yeast sequence cannot sustain replication *instead* of an SV40 AT stretch (lane 8), its presence beside a wild type SV40 *ori* core can inhibit replication (lanes 6 and 7) in a way that reminds us of the AT stretch multimerization experiments and the resulting negative synergism described above (see also lanes 3–5).

**Protein binding properties of the double stranded AT stretch**

What is the reason for the negative synergism observed in *vivo*? And why can the yeast ARS consensus simulate the effect of additional SV40 AT stretches on replication of the SV40 *ori* core? Suspecting that either of the two sequences can be recognized by related (if not the same) factor(s) that down-regulate replication, we analyzed their interactions in *vitro* with CosI cellular factor(s).

Double stranded oligonucleotides of the AT stretch were radiolabelled by substitution of $^{32}$P-dCTP with Klenow enzyme, and used in band shift assays with crude nuclear extract of CosI cells. The data shown in Fig. 5a indicate that binding can be specifically inhibited by a molar excess of non-labelled AT-oligonucleotides (homo AT; Fig. 5a), as well as by the yeast *ori* core consensus (oligo Y), but not by a totally unrelated sequence of similar length (PEB). This implies that there are nuclear factor(s) that can recognize specifically either the SV40 AT stretch or the yeast ARS consensus. The mere presence of polypyrimidine tracts is apparently not enough to grant binding. In fact, competition of binding by 15 bp long $(dA:dT)_{15}$ oligonucleotides (Fig. 5b) is scarcely better than that of $(dG:dC)_{15}$ oligonucleotides.

**Protein binding properties of the two strands of the AT stretch**

We considered that, like other sequences involved in viral regulation (36, 37), and because of their weak base pairing energies, the AT stretches can be bound also when single stranded. Each of the two strands was therefore labelled by kination with $^{32}$P-$\gamma$ATP and used as a probe in gel mobility shift assays with crude nuclear extract of CosI cells. The data indicate that while the A-rich strand ($AT^-$, Fig 6a) provides only weak and unspecific binding, the T-rich strand ($AT^+$) is bound very strongly and specifically (Fig 6b). Here, too, the corresponding single strand from the yeast *ori* core consensus (oligo $Y^+$) was able to compete for binding like $AT^+$, while the unrelated $PEB^+$ oligonucleotide could do so much less efficiently. These data suggest that the AT stretch of the SV40 *ori* core and the yeast ARS consensus can specifically bind to factor(s) both in their double stranded and single stranded forms.

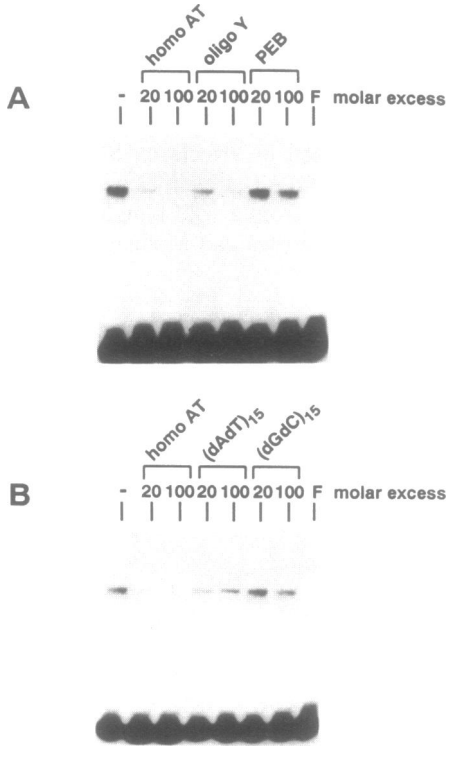


**Fig. 5.** Band shift assays with double stranded AT stretch probes. All samples contain about 8 $\mu$g CosI crude nuclear extracts (see 'Materials and Methods'). A, competition for binding with excess amounts of non-labelled, AT stretch duplex DNA (homo AT), of the homologous sequence from the yeast ARS consensus (oligo Y), and of a totally unrelated sequence (PEB). B, competition for binding with excess amounts of $(dA:dT)_{15}$ and $(dC:dG)_{15}$. The samples without competitors are indicated by a *minus* sign (−). The probe alone (no extract added) is displayed under F. Numbers above the lanes indicate the molar excess ratios used (competitors : probe).

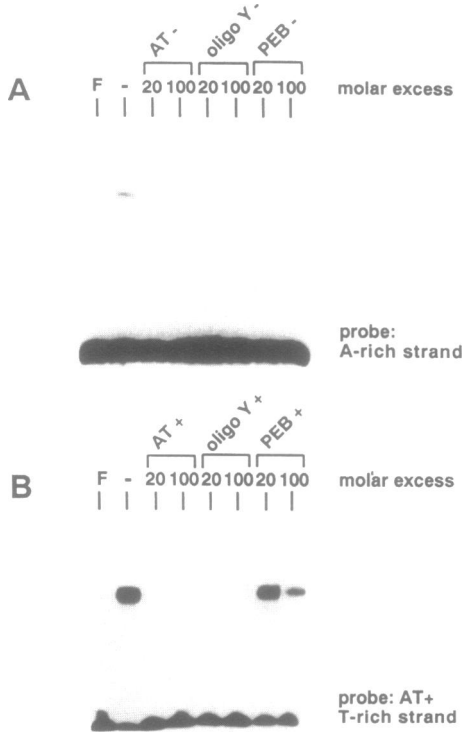


**Fig. 6.** Band shift assays with single stranded AT stretch probes. All samples contain about 8 $\mu$g CosI crude nuclear extracts (see Materials and Methods). A, binding to the A-rich strand ($AT^-$) of the AT stretch; competition for binding with excess amounts of identical but non-labelled single stranded DNA ($AT^-$), of the homologous sequence from the yeast ARS consensus (single-stranded, oligo $Y^-$), and of a totally unrelated, single-stranded sequence ($PEB^-$). B, binding to the T-rich strand ($AT^+$) of the AT stretch; competition for binding with excess amounts of identical but non-labelled single stranded DNA ($AT^+$), of the homologous sequence from the yeast ARS consensus (single-stranded, oligo $Y^+$), and of a totally unrelated sequence (single-stranded, $PEB^+$). The samples without competitors are indicated by a *minus* sign (−). The probes alone (no extract added) are displayed under F. Numbers above the lanes indicate the molar excess ratios used (competitors : probe).

### Binding to single stranded AT stretches is inhibited by double stranded competitors but not *vice versa*

To investigate whether or not there is a relationship between the single and double stranded forms of binding, we started studying the effect of double stranded competitors on binding to single stranded probes . As shown in Fig. 7a, binding to single stranded $AT^+$ can be specifically inhibited by non-labelled, double stranded AT oligonucleotides as well as by single stranded (homo) $AT^+$ oligonucleotides, but not by unrelated double stranded sequences (PEB ds). On the other hand, binding to double stranded AT probes (Fig. 7b) cannot be inhibited by excess amounts of non-labelled, single stranded $AT^+$ oligonucleotides, nor by unrelated single and double stranded sequences. Higher amounts of single stranded competitors (more than 200-fold molar excess) can reduce binding, but then specificity becomes questionable (data not shown).

From these results we conclude that factor(s) bind to the single and double stranded forms of the SV40 *ori* core AT stretch probably in a related manner. However, we cannot yet postulate the identity of these factors.

### Binding to single and double stranded forms of the AT stretch: Southwestern and ultraviolet light-mediated crosslinking assays

To shed light on this matter, we performed a Southwestern blotting assay with about 40 $\mu$g crude nuclear extract of CosI cells. After denaturation and separation through an SDS-polyacrylamide gel, the proteins were electroblotted to a nitrocellulose filter and hybridized with a single stranded (T-rich strand, $AT^+$ SS) and a double stranded (AT DS) radiolabelled probe, respectively. The results are shown in Fig. 8a. It is apparent that in both cases, the main binding species have very similar molecular weights, of about 50 kDa. The same is also

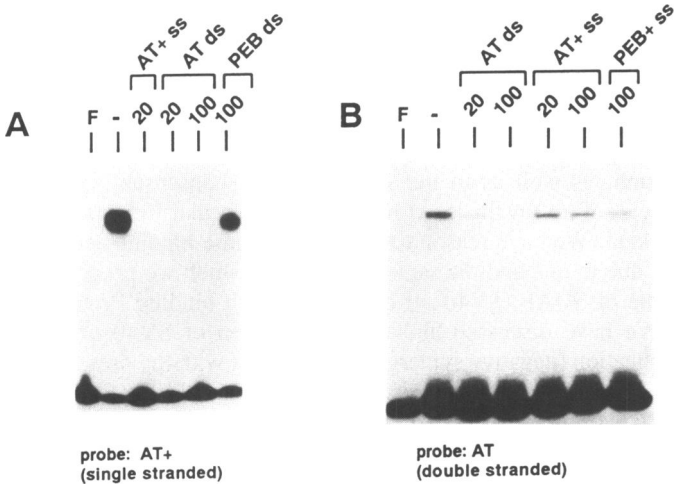


**Fig. 7.** Cross-competition for binding: double vs. single stranded DNA. A, binding to the T-rich strand of the AT stretch; competition for binding with excess amounts of non-labelled, identical DNA ($AT^+$ ss), of the corresponding duplex sequence (AT ds), and of an unrelated sequence (double stranded, PEB ds). B, binding to double stranded probes of the AT stretch; competition for binding with excess amounts of non-labelled, identical DNA (AT ds), of the corresponding single stranded DNA (T-rich strand, $AT^+$ ss), and of an unrelated sequence (single stranded, $PEB^+$ ss). The samples without competitors are indicated by a *minus* sign (−). The probes alone (no extract added) are displayed under F. Numbers above the lanes indicate the molar excess ratios used (competitors : probe).

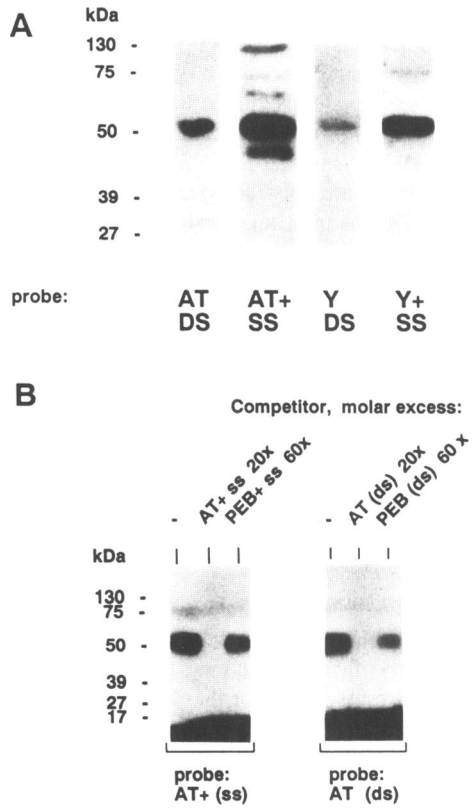


**Fig. 8.** Characterization of the protein species that bind to the SV40 *ori* core AT stretch. A, Southwestern Blotting Assay with about 40 $\mu$g CosI nuclear extract. The assay was performed with double-stranded probes of the SV40 *ori* core AT stretch (AT DS) and of the yeast ARS consensus sequence (Y DS) as well as with the corresponding single, T-rich strands ($AT^+$ SS and $Y^+$ SS). Protein size markers are shown on the left of the panel. B, Ultraviolet light-mediated crosslinking assay. As described in 'Materials and Methods', proteins were allowed to bind to probes of the AT stretch (T-rich strand, $AT^+$ ss; double stranded, AT ds) in absence of competitors (−) or in presence of excess amounts of non-labelled DNA (T-rich strand and unrelated single stranded sequences for the single stranded probe; AT duplex and unrelated duplex DNAs for the double stranded probe).

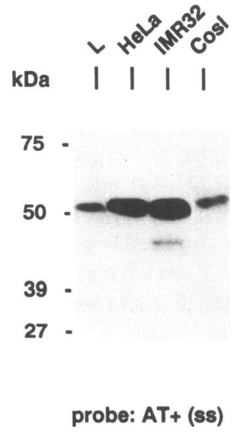


**Fig. 9.** Southwestern blotting with nuclear extracts from different cell lines. About 20 $\mu$g extracts were used. From left to right: mouse fibroblast L cells, human cervical carcinoma HeLa cells, human neuroblastoma IMR32 cells, SV40 Tag-transformed green monkey kidney cells CosI. The probe used is the T-rich strand of the SV40 *ori* core AT stretch.

valid for the yeast ARS consensus, with single stranded (T-rich: $Y^+$ SS) and duplex DNA (Y DS) probes, which corresponds with the findings about competition of binding described above.

In addition, we covalently linked the proteins bound to either single (T-rich strand) or double stranded probes of the AT stretch by irradiation with ultraviolet light (30, 38). Separation through an SDS-polyacrylamide gel yielded signals corresponding, once again, to about 50 kDa molecular weight (Fig. 8b). In accordance with the results obtained above, binding to the single stranded probe can be reduced by excess amounts of the corresponding single stranded homologous DNAs, but not by unrelated sequences. Binding to double stranded probes is reduced by homologous competitors, but to much a lesser extent by unrelated sequences. In the light of these data, we strongly suspect that binding to either forms of the SV40 *ori* core AT stretch—and probably to the yeast ARS consensus—is performed by the same protein, or at least by factors with very similar properties. In such eventuality, we propose the name SOAP, SV40 *ori* core AT stretch binding Protein(s).

### SOAP are ubiquitous factors

We have analyzed the nuclear extracts of different kinds of cells to see whether the factors that bind to the SV40 *ori* core AT stretch (SOAP) are ubiquitous or they are confined to CosI cells. For this purpose, we subjected the extracts to a Southwestern blotting, using the single stranded oligonucleotides ($AT^+$) as probe. Figure 9 displays the results of this experiment: it is apparent that 50 kDa proteins that bind to the SV40 *ori* AT stretch are present in all the cells tested, which have the most disparate provenances. Moreover, comparative band shift assays with extracts from CosI and human leukemia HL60 cells (using $AT^+$ probe) yielded identical patterns with respect to specificity of binding and migration of the nucleoprotein complexes (data not shown). We therefore believe that SOAP are ubiquitous factors.

## DISCUSSION

The AT-rich stretch of the SV40 origin of replication core sequence is known to be necessary for replication *in vivo* (1, 2, 4, 5) and *in vitro* (2). Replication activity is lost upon mutations, deletions as well as small insertions (3, 14), and changing of the relative orientation to the rest of the *ori* core (as demonstrated by some of the data presented here). These reports suggest that this sequence regulates replication by influencing DNA bending, melting and/or unwinding of the SV40 *ori* core. On the other hand, several groups report the identification of several proteins that bind to the AT-rich tract (19) and that are probably involved in replication of SV40 DNAs: one of them was even isolated from the multiprotein DNA polymerase complex of HeLa cells (20). In addition, it has also been reported that the region of SV40 DNA comprising the AT stretch and part of Auxiliary Sequence II (overlapping with the GC-rich repeats that facilitate initiation) is the binding site for factor(s) necessary for initiation of replication (39).

So far, no one had ever analyzed the effect of multimerizing the AT stretches on SV40 *ori* core replication. Considering that multimerization of binding sites for transcriptional factors often results in increased transcription (21, and references therein), potential synergism was an interesting topic to study. In this work we have found that, surprisingly, the presence of additional AT stretches in *cis* beside a functional *ori* core greatly reduces its replication. Two explanations can be offered: (1) either the presence of more AT stretches alters the structure of the DNA in such a way that replication is inhibited; or (2) the presence of additional binding sites for cellular factors creates a sort of physical block that interferes with the normal course of replication. The two hypotheses need not be mutually exclusive. Moreover, a similar inhibition of replication can be induced by the presence of a sequence reminiscent of the SV40 AT stretch, the yeast ARS consensus (33–35, 40): it inhibits replication in *vivo* when added in *cis* to a functional SV40 *ori* core, just like additional wild type SV40 AT stretches. However, it cannot sustain replication *instead* of the wild type AT stretch within an SV40 *ori* core. The most likely interpretation is that those cellular factors which inhibit replication when binding to supplementary SV40 AT stretches can also recognize the yeast ARS consensus.

This view is supported by the data obtained in *vitro*, where we have investigated the binding properties of crude CosI nuclear extract to the AT stretch. Its A-rich strand is bound only weakly and non-specifically. The duplex form and the T-rich strand, instead, are bound sequence-specifically. The yeast ARS consensus, homologous to the SV40 AT stretch, can compete with the latter for binding, while totally unrelated sequences (as well as sequences that like $(dA:dT)_{15}$ are structurally different) are unable to do so. This implies that the same factors recognize specifically either the SV40 AT stretch or the yeast ARS consensus, and we can all but speculate that we are observing here the mammalian counterpart to the ARS binding activity of yeast cells. Incidentally, in the yeast origins of replication, too, the T-rich strands are bound preferentially (41–42). Other reports also demonstrate the role of proteins that bind to single stranded targets in a variety of viral and cellular processes (36, 37, 43–50). Further, we have found that binding to the T-rich strand is reduced sequence-specifically by a molar excess of the corresponding double stranded DNA. By contrast, inhibition of binding to duplex DNA by single stranded AT stretches is possible only with high amounts of competitors and with very dubious specificity. These data strongly suggest that binding to the single and double stranded forms of the SV40 *ori* core AT stretch are tightly connected. Results from Southwestern and ultraviolet light-mediated crosslinking assays indicate that specific binding to both single and double stranded forms of the AT stretch, as well as to the yeast ARS *ori* consensus, involves species of (to say the least) very similar molecular weights, about 50 kDa. We have reason to believe that these binding activities are due to one and the same protein, for which we propose the name of SOAP, SV40 *ori* core AT stretch binding Protein(s).

We have discussed above that inhibition of SV40 *ori* core replication (negative symergism) correlates with the presence in *cis* of additional binding sites for SOAP, namely the SV40 AT stretch or the yeast ARS consensus. What are the biological functions of SOAP ? One may think that, like most other factors binding to AT-rich sequences (20, 39, 51), SOAP belongs to the replication machinery, and that it may work alternatively as an inhibitor in the presence of excess binding sites. In this case, one copy of the yeast ARS consensus *instead* of the SV40 AT stretch would be expected to sustain replication, but that is not the case (Fig.4, lane 8). A likely explanation is that while inhibition may require only protein binding, replication initiation requires in addition a very precise DNA structure. On top of being a target for cellular factors, the SV40 AT stretch favours replication by forcing DNA to assume a particular conformation (52), while the yeast sequence may be unable to do so.

Alternatively, although replication factors can bind to the yeast ARS, its particular sequence may prevent them to perform their biological duties. On the other hand, of course, we cannot exclude *a priori* the possibility that SOAP's main role is to downregulate replication, and that replication-promoting activity resides in other AT stretch-binding proteins that do not interact with the yeast ARS consensus and that are not identified here. This last hypothesis, however, somehow conflicts with the fact that binding of SOAP to the SV40 AT stretch is dominant: if SOAP were an inhibition-only factor, even a wild type SV40 *ori* core would not replicate. In order to study more in detail the actual functional role of SOAP, as well as its biochemical characters, we envisage its cloning and purification.

As mentioned at the beginning of this discussion, the observation *per se* that cellular factors bind to adenine and thymine-rich sequences is not new. At first, one wonders why the binding activity to the SV40 AT stretch which we have imputed to SOAP had not been detected earlier. The fact is that all other proteins which specifically interact with AT-rich sequences were either purified from isolated protein complexes belonging to the replication machinery (20, 51), or their existence (but not their character) was assessed with more genetical methods (39). To our knowledge, this is the first direct biochemical approach that makes use of crude nuclear extracts in combination with SV40 AT stretch oligonucleotides. With this regard, SOAP represents a new category of factors, distinct from the proteins mentioned above (19, 20, 39, 51). Not only its replication-inhibiting function distinguishes it, but it also differs from them in binding affinities, specificities, strand preference, and molecular weight.

## ACKNOWLEDGEMENTS

We thank Ms. Mako Yamagishi for accurate secretarial work and Mrs. Kiyomi Takaya for skillful technical help. We are also indebted to Drs. Walter Schaffner, Sandro Rusconi, Edgar Schreiber and Yasuhiro Imamura for helpful discussions. This work was supported by grants from the Ministry of Education, Science and Culture in Japan, Nissan Science Foundation, Akiyama Foundation, Suhara Memorial Foundation and Hayashi Memorial Foundation for Female Natural Scientists. I.G. is receiving a Foreign Student Scholarship from the Japanese Government, Ministry of Education (*Kokuhi Ryugakusei*).

## REFERENCES

1. Gerard,R. and Gluzman,Y. (1985) *EMBO J.*, **4**, 2933–2939.
2. Stillman,B., Gerard,R.D., Guggenheimer,R.A. and Gluzman,Y. (1985) *EMBO J.*, **4**, 2933–2939.
3. Deb,S., De Lucia,A.L., Baur,C.-P. Koff,A. and Tegtmeyer,P. (1986) *Mol. Cell. Biol.*, **6**, 1663–1670.
4. Li,J.J., Peden,K.W.C., Dixon,R.A.F. and Kelly,T. (1986) *Mol. Cell. Biol.*, **6**, 1117–1128.
5. Hertz,G.Z., Young,M.R. and Hertz,J.E. (1987) *J. Virol.*, **61**, 2322–2525.
6. De Lucia,A.L., Deb,S., Parlin,K. and Tegtmeyer,P. (1986) *J. Virol.*, **57**, 138–144.
7. Guo,Z.S., Gutierrez,C., Heine,U., Sogo,J.M. and De Pamphilis,M.L. (1989) *Mol. Cell. Biol.*, **9**, 3593–3602.
8. Lee-Chen,G.J. and Woodsworth Gutai,M. (1986) *Mol. Cell. Biol.*, **6**, 3086–3093.
9. Gutierrez,C., Guo,Z.S., Roberts,J. and De Pamphilis,M.L. (1990) *Mol. Cell. Biol.* **10**, 1719–1728.
10. Hay,R.T. and De Pamphilis,M.L. (1982) *Cell*, **28**, 767–779.
11. Hay,R.T., Hendrickson,E.A. and De Pamphilis,M.L. (1984) *J. Mol. Biol.*, **175**, 131–157.
12. De Pamphilis,M.L. and Bradley,M.K. (1986) In Salzman,N.P. (ed.),*The Papoviridae*, Plenum Pub. Corp., Vol. I, 99–243.
13. Banerji,J., Rusconi,S. and Schaffner,W. (1981) *Cell* , **27**, 299–308.
14. Deb, S., De Lucia,A.L., Koff,A., Tsui,S. and Tegtmeyer,P. (1986) *Mol. Cell. Biol.*, **6**, 4578–4584.
15. Tamanoi,F. and Stillman,B.W. (1983) *Curr. Top. Microbiol. Immunol.*, **109**, 75–87.
16. Kearsey,S. (1984) *Cell*, **37**, 299–307.
17. Weller,S.A., Spadaro,A., Schaffer,J.E., Murray,A.W., Maxam,AM. and Schaffer P.A. (1985) *Mol. Cell. Biol.*, **5**, 930–942.
18. Zahn,K. and Blattner,F.R. (1985) *Nature*, **317**, 451–453.
19. Baur,C.-P. and Knippers,R. (1988) *J. Mol. Biol.*, **203**, 1009–1019.
20. Malkas,L.H. and Baril,E.F. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 70–74.
21. Schatt,M.D., Rusconi,S. and Schaffner,W. (1990) *EMBO J.*, **9**, 481–487.
22. Graham,F.L. and van der Eb,A.J. (1973) *Virology*, **52**, 456–467.
23. Hirt,J.B. (1967) *J. Mol. Biol.*, **26**, 365–369.
24. Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503–517.
25. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
26. Dignam,J.D., Lebovitz,R.M. and Roeder,R.G. (1983) *Nucl. Acids Res.*, **11**, 1475–1489.
27. Westin,G., Gerster,T., Müller,M.M., Schaffner,G. and Schaffner,W. (1987) *Nucl. Acids Res.*, **15**, 6787–6797.
28. Fried,M. and Crothers,D.M. (1981) *Nucl. Acids Res.*, **9**, 6501–6525.
29. Garner,M.M. and Revzin,A. (1981) *Nucl. Acids Res.*, **9**, 3047–3060
30. Chodosh,L.A., Carthew,R.W. and Sharp,P.A. (1986) *Mol. Cell. Biol.*, **6**, 4273–4733.
31. Laemmli,U.K. (1970) *Nature*, **227**, 680–685.
32. Miskimins,W.K., Roberts,M.P., McClelland,A. and Ruddle,F.H. (1985) *Proc.Natl. Acad. Sci. USA*, **82**, 6741–6744.
33. Stinchcomb,D.T., Mann,C., Setker,L. and Davis,R.W. (1981) *ICN–UCLA Symp. Mol. Cell. Biol.*, **22**, 473–488.
34. Broach,J.R., Li,Y.Y., Feldman,J., Jayaram,M. Abraham,K., Nasmyth,A. and Hicks,J.B. (1983) *Cold Spring Harbor Symp. Quant.Biol.*, **47**, 1165–1173.
35. Van Houten,J.V. and Newlon,C.S. (1990) *Mol. Cell. Biol.*, **10**, 3917–3925.
36. Gaillard,C. and Strauss,F. (1990) *J. Mol. Biol.*, **215**, 245–255.
37. Finocchiaro,L.M.E., Amati,P. and Glikin,G.C. (1991) *Nucl. Acids Res.*, **19**, 4279–4287.
38. Negishi,Y., Iguchi-Ariga,S.M.M. and Ariga,H. (1992) *Oncogene*, **7**, 543–548.
39. Yamaguchi,M. and De Pamphilis,M.L. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 1646–1650.
40. Tschumper,G. and Carbon,J. (1980) *Gene*, **50**, 157–166.
41. Hofmann,J.F.X. and Gasser,S.M. (1991) *Cell* , **64**, 951–960.
42. Schmidt,A.M.A., Herterich,S.M. and Krauss,G. (1991) *EMBO J.*, **10**, 981–985.
43. Traut,W. and Fanning,E. (1988) *Mol. Cell. Biol.*, **8**, 903–911.
44. Roller,R.J., McCormick,L. and Roizman,B. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 6518–6522.
45. Mukherjee,R. and Chambon,P. (1990) *Nucl. Acids Res.*, **18**, 5713–5716.
46. Wilkinson,W.O., Min,H.Y., Claffey,K.P., Satterberg,B.L. and Spiegelman,B.M. (1990) *J. Biol. Chem.*, **265**, 477–482.
47. Bergemann,A.D. and Johnson,E.M. (1992) *Mol. Cell. Biol.*, **12**, 1257–1265.
48. Jansen-Dürr,P., Boshart,M., Lupp,B., Bosserhoff,A., Frank,R.W. and Schütz, G. (1992) *Nucl. Acids Res.*, **20**, 1243–1249.
49. Kolluri,R., Torrey,T.A. and Kinniburgh,A.J. (1992) *Nucl. Acids Res.*, **20**, 111–116.
50. McKay,S.J. and Cooke,H. (1992) *Nucl. Acids Res.*, **20**, 1387–1391.
51. Fry,M., Lapidot,J. and Weisman-Shomer,P. (1985) *Biochemistry*, **24**, 7549–7556.
52. Travers,A.A. (1990) *Cell*, **60**, 177–180.