JOURNAL OF VIROLOGY, June 1988, p. 1999–2006
0022-538X/88/061999-08$02.00/0

Vol. 62, No. 6

Expression of Simian Virus 40 T Antigen in *Escherichia coli*: Localization of T-Antigen Origin DNA-Binding Domain to within 129 Amino Acids

AVRIL K. ARTHUR, ADOLF HÖSS, AND ELLEN FANNING*

Institute for Biochemistry, Karlstrasse 23, 8000 Munich 2, Federal Republic of Germany

Received 25 August 1987/Accepted 18 February 1988

The genomic coding sequence of the large T antigen of simian virus 40 (SV40) was cloned into an *Escherichia coli* expression vector by joining new restriction sites, *Bgl*II and *Bam*HI, introduced at the intron boundaries of the gene. Full-length large T antigen, as well as deletion and amino acid substitution mutants, were inducibly expressed from the *lac* promoter of pUC9, albeit with different efficiencies and protein stabilities. Specific interaction with SV40 origin DNA was detected for full-length T antigen and certain mutants. Deletion mutants lacking T-antigen residues 1 to 130 and 260 to 708 retained specific origin-binding activity, demonstrating that the region between residues 131 and 259 must carry the essential binding domain for DNA-binding sites I and II. A sequence between residues 302 and 320 homologous to a metal-binding "finger" motif is therefore not required for origin-specific binding. However, substitution of serine for either of two cysteine residues in this motif caused a dramatic decrease in origin DNA-binding activity. This region, as well as other regions of the full-length protein, may thus be involved in stabilizing the DNA-binding domain and altering its preference for binding to site I or site II DNA.

The large T antigen of simian virus 40 (SV40) is a particularly interesting model protein for the study of the control of eucaryotic transcription, DNA replication, and cell growth (reviewed in references 6 and 34). Repression of early viral transcription is dependent on specific DNA binding of T antigen to two major sites in the SV40 origin of replication (7, 14, 35). Origin-specific DNA-binding activity of T antigen may also be required, perhaps indirectly, for the *trans*-activation of late viral gene expression (3, 17) and the regulation of cellular gene expression (40, 41, 46). Replication of viral DNA also requires specific binding of T antigen to the origin of replication, in particular to site II (28, 42, 47).

The two major T-antigen-binding sites in the SV40 origin of replication have been studied in great detail (5, 16, 50, 51). Site I harbors two tandem repeats of the pentanucleotide 5′-GAGGC, separated by an oligo(dT) sequence and a third, nonessential, imperfect pentanucleotide. Site II bears two pairs of 5′-GAGGC repeats arranged in a palindrome. Each pentanucleotide directs the binding of one monomer mass of T antigen, which appears to recognize and interact with this binding signal (24, 37).

However, a description of the precise protein-DNA interactions of T antigen with these two different sites at a level comparable to that of procaryotic DNA-binding proteins (33) requires more detailed knowledge of the minimal peptide sequence required for specific origin DNA binding. Attempts to define the origin DNA-binding domain of T antigen have often been based on loss of function and have yielded contradictory results. Analysis of truncated and chimeric T-antigen polypeptides indicated that the origin-binding domain was localized in the amino-terminal half of the second exon (4, 31). Pseudorevertants encoding T antigens able to phenotypically suppress several different point mutations in binding sites I and II carried mutations that also mapped to this region (23). Multiple point mutations or small deletions that disrupted origin-binding activity mapped to this region but also to sequences outside this region (4, 22, 29, 32, 49). Limited proteolysis of T antigen was used to identify a peptide comprising residues 1 to 130 that retained origin-binding function (27), but a more recent report implicated a peptide containing residues 131 to 371 (44). Interestingly, this larger peptide carries a sequence homologous to a single metal-binding "finger," implicated in the DNA-binding specificity of transcription factor (TF) IIIA (25) and also found in other eucaryotic proteins involved in transcriptional control (reviewed in references 2 and 9).

In the present report, we analyze the origin DNA-binding activity of wild-type and mutant T-antigen polypeptides expressed in *Escherichia coli*. We demonstrate that T antigen can be produced in *E. coli* in significant amounts and that it specifically binds to DNA sequences in the SV40 origin. Point mutations in the finger motif inactivate origin DNA binding. However, truncated T antigens lacking this region or the amino terminus retain origin-specific DNA-binding activity. The results suggest that the origin DNA-binding region of T antigen is located between amino acids 131 and 259 but that the presence of additional sequences may help to stabilize this region or alter the preference of DNA binding to sites I or II.

(A preliminary report on this work was presented at the DNA Tumor Virus Meeting on SV40, Polyoma, and Adenoviruses in Cold Spring Harbor, N.Y., 13–17 August 1986 [abstr. no. 103, p. 103].)

MATERIALS AND METHODS

Plasmid construction. Cloning procedures and miniscale plasmid analysis followed routine procedures (21); enzymes were used as recommended by the manufacturer (Pharmacia, Bethesda Research Laboratories, and New England Biolabs). Restriction fragments were isolated from agarose gels on DEAE paper (NA45; Schleicher & Schüll) as described (8). Site-directed oligonucleotide mutagenesis was done with the M13 origin-containing plasmid pSDL13 grown in *E. coli* XS127 (20) with the entire SV40 genome cloned in

* Corresponding author.

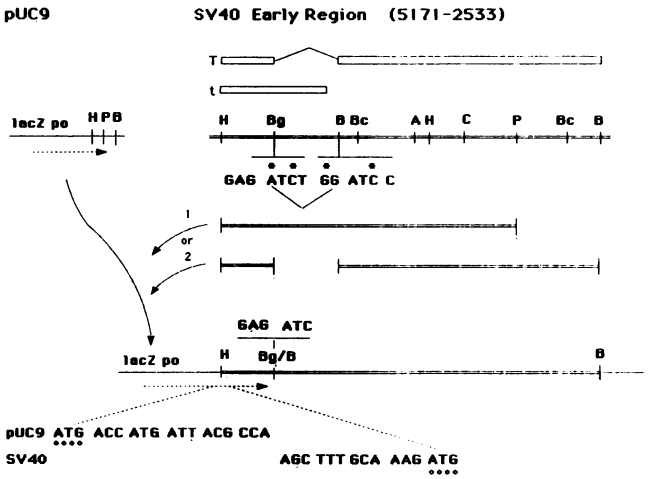

FIG. 1. Cloning scheme. The new *Bgl*II (Bg) and *Bam*HI (B) restriction sites were introduced into the early region of SV40 to create the base pair changes indicated (*). After *Bgl*II-*Bam*HI ligation, only a single silent change occurred in the codon for Ile-83; ATT → ATC. By cloning the wild-type SV40 *Hin*dIII-*Pst*I (nucleotides 5171 to 3204) fragment into the polylinker of pUC9 as depicted in the first cloning, the coding region of small t antigen was retained. After cloning the fragments *Hin*dIII-*Bgl*II (5171 to 4914), *Bam*HI-*Pst*I (4568 to 3204), and *Pst*I-*Bam*HI (3204 to 2533) as depicted in the second cloning, the result is an intronless copy of the T antigen, as shown below with the DNA sequence after the *Bgl*II-*Bam*HI ligation. The predicted N-terminal sequence of the T antigen is shown below. Other restriction sites used in cloning were *Hin*cII (C; 3733), *Hin*dIII (H; 4002), *Sau*3A (A; 4100), and *Bcl*I (Bc; 2770). A novel *Bcl*I site was created at position 4427.

the *Bam*HI site of the polylinker; the mutagenesis procedures are described elsewhere (39, 56). The oligonucleotides were provided by R. Mertz and D. Weigand (Genzentrum, Munich).

Bacterial growth conditions. *E. coli* JM103 was the host in all protein expression experiments shown. A strain with reduced protease activity, SG935IQ (*htpR*, *lon* [12]) (gift of Robert Garcea) was also tested but found not to yield significantly different results. Since the plasmids expressing T-antigen sequences were relatively unstable or maintained at a reduced copy number in *E. coli*, every extract was prepared from freshly transformed JM103 colonies picked while quite small. Plasmid-containing bacteria survived longer when propagated in minimal medium plus glucose rather than L-broth. The antibiotic cephaloridine (20 μg/ml) gave better selection than ampicillin. Overnight cultures were induced by diluting 1:25 in L-broth minus glucose, growing at 37°C for 1 h before addition of isopropyl-β-D-thiogalactopyranoside (IPTG) (5 mM), and further incubation at 37°C for 3 h.

Cell extracts and immunoprecipitation of T antigen. Induced cultures of *E. coli* (10^8 cells per ml) were pelleted and suspended directly in Laemmli sample buffer (19), heated to 95°C for 15 min, and spun, and the supernatant was loaded on polyacrylamide gels. Alternatively, cells were suspended in lysis buffer (1/50 of the culture volume; 50 mM Tris [pH 9], 120 mM NaCl, 0.5% Nonidet P-40, 5% glycerol, 5 mM EDTA, 5 mg of lysozyme per ml, 1 mM dithiothreitol [DTT], 1 mM phenylmethylsulfonyl fluoride [PMSF]) and kept at 0°C for 1 h before sonication (three times, 15 s each). The supernatant of spun lysates was divided into portions and frozen at −70°C; freshly thawed portions were used for every experiment and then discarded. T-antigen proteins were immunoprecipitated by adding 5 μg of purified monoclonal antibody to appropriate amounts of *E. coli* extract plus lysis buffer in a final volume of 100 μl, essentially as described (10, 54). After washing, immune complexes were suspended in either Laemmli sample buffer for Western blot (immunoblot) analysis or DNA-binding buffer.

Immunoblots. Transfer of proteins from sodium dodecyl sulfate (SDS)-polyacrylamide gels (21) to nitrocellulose paper was performed as described earlier (52). Antibody staining was carried out as described earlier (54), except that an alkaline phosphatase-conjugated second antibody was used (Promega Biotec).

Origin DNA-binding assays. Wild-type origin-containing plasmid pSVwt carries the entire SV40 genome (11); site I-containing plasmid pONwt has a 19-base-pair (bp) synthetic oligonucleotide of site I cloned into pAT153 (37); and the site II-containing plasmid p1097 was the entire SV40 genome of mutant cs1097 (7), which carries a 31-bp deletion encompassing all of site I, cloned in pAT153. Plasmids were digested (pSVwt, *Hin*dIII; pONwt, *Hin*dIII, *Sal*I; p1097, *Hin*dIII) to generate suitable fragments, dephosphorylated, and end-labeled with polynucleotide kinase and [γ-^{32}P]ATP as described (11). Excess DNA was added to antibody-bound immunopurified T antigen (described above) which had been incubated in DNA-binding buffer (10 mM HEPES [*N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid, pH 7.8], 80 mM KCl, 1 mM DTT, 1 mM PMSF, 0.2 mg of glycogen per ml, 1 mg of bovine serum albumin per ml at 0°C for 1 h, and incubation was continued at 0°C for 2 h. Bound DNA was dissociated from immune complexes and analyzed by 1.5% agarose gel electrophoresis and autoradiography as described (11, 36).

Immunoaffinity chromatography. An induced 100-ml culture of JM103 carrying pTh was collected by centrifugation. Cells were suspended on ice in 2.5 ml of buffer A (50 mM Tris [pH 8.0], 50 mM NaCl, 1 mM EDTA, 1 mM PMSF, 1 mM DTT, 10% glycerol), treated with lysozyme (5 mg/ml) for 15 min, sonicated, and incubated for 30 min on ice. T antigen was adsorbed to 2.5 ml of Q-Sepharose FF slurry (Pharmacia) equilibrated in buffer A. After centrifugation, the supernatant was discarded and T antigen was eluted in 2.5 ml of buffer B (50 mM Tris [pH 8.0], 250 mM NaCl, 1 mM EDTA, 1 mM PMSF, 1 mM DTT, 10% glycerol). Further purification of T antigen by monoclonal antibody (PAb) 419-protein A-Sepharose chromatography followed published protocols (43). T-antigen-containing fractions, detected by Western blot, were combined and further purified by fast-protein liquid chromatography (FPLC) Mono Q chromatography (Pharmacia).

RESULTS

Construction of plasmids for T-antigen expression in *E. coli*. To create an intronless copy of the coding sequence of the large T antigen of SV40, a *Bgl*II site and *Bam*HI site were introduced at the 5′ and 3′ intron boundaries, respectively, by oligonucleotide-directed mutagenesis. Ligation of the complementary *Bgl*II and *Bam*HI ends generated an uninterrupted T-antigen coding sequence with no amino acid changes (Fig. 1). Other nucleotide changes introduced by mutagenesis created either stop codons (T-131, T-143, T-260), amino acid changes (Cys-302 → Ser; Cys-305 → Ser), or a novel *Bcl*I site (Fig. 1 and 2). Appropriate DNA fragments were cloned into the polylinker of pUC9; a simple example of this strategy, the cloning of small t antigen or full-length T antigen, is shown in Fig. 1.

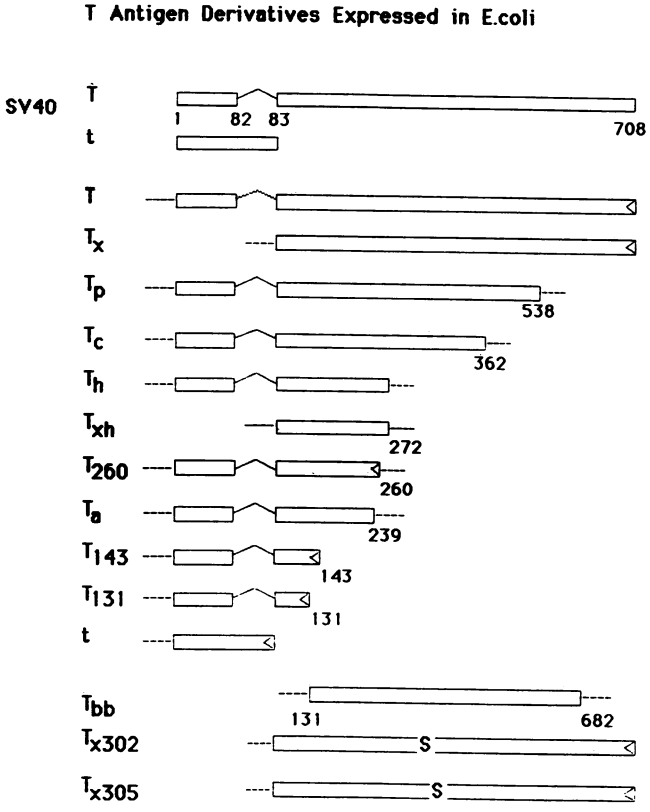

FIG. 2. Diagram of mutant T-antigen constructs cloned and expressed in *E. coli*. Above is shown the early region of SV40 with the two translation products of the large and small T antigens. Below are the translational products predicted after cloning into pUC9. The nomenclature indicates the derivation of the construct: x, second exon; p, c, h, a, and b, restriction sites *Pst*I, *Hin*cII, *Hin*dIII, *Sau*3A, and *Bcl*I, respectively, used to truncate the C terminus or both N and C termini; codon numbers 260, 143, and 131 indicate new stop codon positions; 302 and 305 indicate substitution of Ser for Cys at codons 302 and 305, respectively. N-terminal and C-terminal fusion residues are marked (– – –). In Tx302 and Tx305, the Ser substitutions for Cys-302 and Cys-305 are indicated (S). The codon number of the stop codons (◁) and the last SV40-coded T-antigen residue number are indicated for each construct. Tx and Txh have 12 and Tbb has 13 N-terminal fusion residues; Tc, Th, Txh, Ta, and Tbb carry 18 C-terminal fusion residues; Tp encodes a predicted 55 C-terminal fusion residues (53, 55).

Cloning SV40 sequences into the *Hin*dIII site of the pUC9 polylinker, located 15 bp 3′ of the *lacZ* AUG start codon, generates an in-frame readthrough from the *Hin*dIII site at nucleotide 5171 of SV40 into the AUG start of t/T antigen located 9 bp downstream. Thus, all clones encoding the 5′ end of the t/T antigen contained a postulated nine N-terminal fusion residues (Fig. 1 and 2). Other N-terminal and C-terminal fusion residues resulting from construction of mutants are indicated in Fig. 2.

Expression of t/T antigen derivatives in *E. coli*. The specific t/T antigen residues encoded by each construct (Fig. 2; see legend for explanation of construct nomenclature) were confirmed by analysis of the size and monoclonal antibody epitopes presenting in the polypeptides by Western blots (Fig. 3 and 4). In all cases the largest polypeptides had the size expected for the full-length gene product. Some of the constructs also produced degradation products. To facilitate comparison of the amounts of T-antigen-specific polypeptides produced, the volume of induced *E. coli* culture applied to the gels for each construct is listed in the legend to Fig. 4.

The longest polypeptides, particularly full-length T antigen, were poorly expressed. Since overloading acrylamide gels led to band distortion, these proteins were best detected by Western blots after immunoprecipitation of lysed *E. coli* extracts. Immunoprecipitated full-length T antigen expressed in *E. coli* (Fig. 3, lanes 3 and 5) showed a major protein species of the same size as the eucaryotic T antigen from COS 1 cells (lanes 2 and 4). Tx, Tx302, and Tx305 were detected at equal levels after immunoprecipitation by either Pab 416 (Fig. 3, lanes 6, 7, and 9) or the denaturation-sensitive Pab 1630 (lanes 8 and 10) and Western blotting with Pab 416. As predicted, all three peptides were of identical size; none of them reacted with Pab 108 (Fig. 3).

Small t antigen, T131, T143, Ta, T260, and Th all interacted with the N-terminal specific antibodies Pab 108 (Fig. 4A, lanes 4 to 8, 13, and 14) and Pab 419 (not shown), but as predicted, Txh (lane 9) did not, since the only proteins detected by the antibodies were also present in the control (lane 10). However, Pab 416 did detect a specific Txh peptide (lane 11) not detected in the control extract (lane 12), as well as the T131 and T143 peptides (Fig. 4A, lanes 2 and 3). Small-t-antigen extract did not react with Pab 416 (lane 1). These results confirmed the expression of predicted sequences by each construct. In addition, the epitope of Pab 416 was more precisely mapped to a sequence between T-antigen residues 83 and 130 (15).

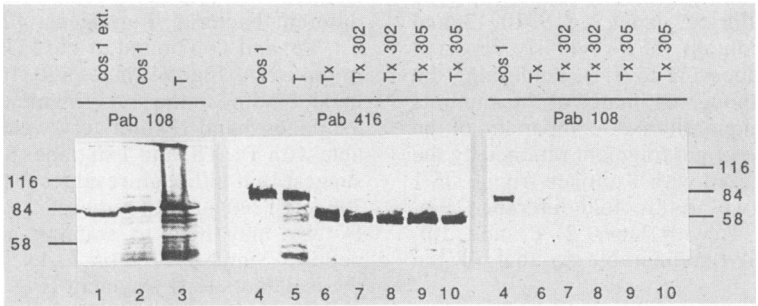

FIG. 3. Western blots of immunoprecipitated *E. coli* extracts. Whole-cell extracts were prepared from the indicated cultures and immunoprecipitated with Pab 108, which recognizes an epitope in the N-terminus of t/T antigen (13) (lanes 2 and 3), Pab 416, which recognizes an epitope in the N-terminal end of the second exon of T antigen (15) (lanes 4 to 7 and 9), and Pab 1630, which recognizes a denaturation-sensitive epitope in the second exon (1) (lanes 8 and 10). T antigen was detected in Western blots with Pab 108 or Pab 416, as indicated. Extract from COS 1 cells (cos 1 ext.) and immunoprecipitated COS 1 T antigen served as controls (lanes 1 and 2). The migration of marker proteins is shown on the left and right (in kilodaltons). Lanes 1 to 3, 13.75% PAGE, lanes 4 to 10, 12% PAGE.

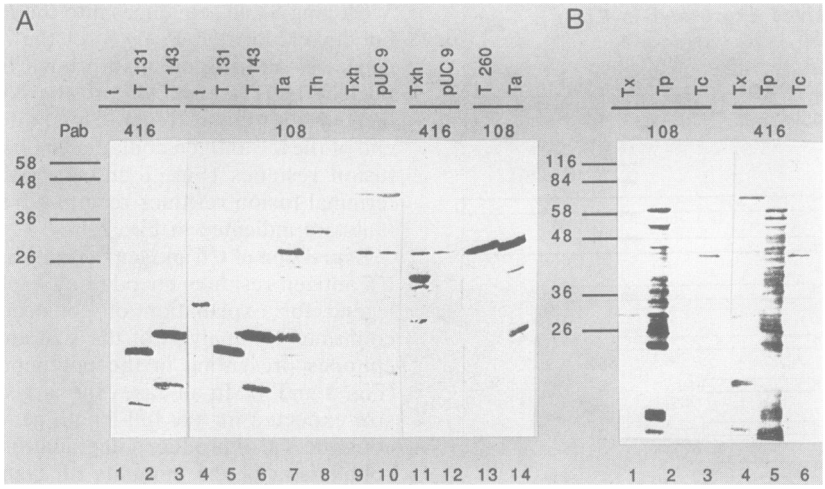

FIG. 4. Western blots of proteins expressed by different mutants in *E. coli* after induction with IPTG. Whole cells from the indicated volumes of induced cultures (10^8 cells per ml) were boiled in SDS sample buffer and electrophoresed in 15% (A) and 13.75% (B) polyacrylamide gels. Extract from a pUC9-bearing culture served as a control. Blots were treated as indicated with Pab 108. IgG and Pab 416 IgG. The migration of marker proteins is shown at the left (in kilodaltons). (A) Lanes 1, 4, 6, and 9 to 14, 0.1 ml; lane 7, 0.05 ml; lanes 3, 6, and 8, 0.01 ml; lanes 2 and 5, 0.005 ml. (B) Lanes 1 and 4, 0.25 ml; lanes 2 and 5, 0.1 ml; lanes 3 and 6, 0.025 ml.

Similarly, the larger constructs Tx, Tp, and Tc all interacted with Pab 416 (Fig. 4B, lanes 4 to 6), while only Tp and Tc were also detected by Pab 108 (lanes 1 to 3). In contrast to the other constructs, Tp, which contained the largest C-terminal fusion, was highly degraded (Fig. 4B, lanes 2 and 5). Tbb was not detected by Western blot, since none of the antibodies recognized epitopes in this region (residues 131 to 682).

Binding of T antigen to SV40 origin DNA. Approximately equivalent amounts of T-antigen derivatives, as estimated by immunoblot, were immunoprecipitated with an appropriate antibody (Pab 108, 419, 416, or 1630). Since the immunoprecipitate was washed extensively before the addition of end-labeled DNA, this served to remove other *E. coli* proteins and unbound T antigen species from the sample. Binding of full-length and mutant T antigens to SV40 DNA was tested with three different templates: one containing the wild-type origin sequences, one containing only site I, and one containing only site II (see Materials and Methods) (Fig. 5).

When binding experiments were performed with intact SV40 origin DNA, alone (not shown) or as an equimolar mixture with site I (Fig. 5A), full-length T antigen and mutants Tp, Tx, and Tbb bound specifically to the wild-type origin-containing fragment (Fig. 5A, lanes 2, 8, 9, 10, 13, and 14). These constructs all contain the previously reported origin-binding peptide (residues 131 to 371), confirming the results of Simmons (44). Although estimates of the amounts of each protein are only semiquantitative, comparison of the relative amounts of wild-type origin fragment retained by the larger proteins (T, Tx) compared with T antigen from COS 1 cells revealed at the most a two- to threefold difference (Fig. 5A, compare lanes 1 and 7 with lanes 2, 8, and 10), suggesting a binding activity comparable to that of the eucaryotic protein.

The mutants Tx302 and Tx305 are identical to Tx, apart from a Cys → Ser substitution at residue 302 or 305 in a sequence homologous to a zinc finger motif (2). These T antigens did not bind specifically to the SV40 origin fragment (Fig. 5A, lanes 15 and 16) or to the isolated site I (Fig. 5A) or site II templates (Fig. 5B, lanes 7 and 8). This result indicated that the finger motif may be important for origin DNA-binding activity, but it is not clear whether it is directly involved in DNA binding or simply disrupts the native protein conformation.

Therefore, mutants carrying more extended deletions of T-antigen coding sequences were assayed for origin DNA-binding activity. Th and Txh, which lack T antigen sequences beyond residue 272, and T260, which carries a stop codon at position 260, bound specifically to intact SV40 origin DNA (Fig. 5A, lanes 4, 5, 11, 12, and 17). In contrast, the mutants T131, T143, small t antigen (data not shown), and Ta (T antigen residues 1 to 239) did not bind specifically to origin DNA (Fig. 5A, lanes 6 and 18).

As a control, the region corresponding to residues 83 to 272 from Tx302 and Tx305 was recloned to create derivatives equivalent to Th. Both of the truncated proteins regained DNA-binding ability with a binding pattern identical to that of Th (not shown), confirming that the origin DNA-binding region of each mutant was intact. Thus, DNA sequences encoding T-antigen residues 1 to 130 and 260 to 708 could be deleted from the construct without eliminating origin DNA binding.

Differences in relative binding to intact origin DNA and isolated site I and site II templates were observed among the different bacterial T antigens. For example, full-length T, Tx, Tp, and Tbb bound to site I DNA just as well as to intact origin DNA (Fig. 5A, lanes 8, 9, 10, 13, and 14). On the other hand, binding to the site I fragment was weak with T260 (Fig. 5A, lanes 5 and 12) and very weak (not shown) or undetectable with Tc, Th, and Txh (lanes 3, 4, 11, and 17). These data suggest that although residues 1 to 130 (cf. Tbb) and 260 to 708 (cf. T260) are not required for binding to the isolated site I, they may help to stabilize the binding region of the peptide. Similarly, *E. coli* T, Tx, Th, and Txh bound well to the isolated site II fragment (Fig. 5B, lanes 1, 3, 11, 12, 13, and 14), while Tbb and T260 bound only weakly (lanes 4 and 6). Again, the data suggest that residues 1 to 130 and 260 to 708 may be deleted without eliminating site II binding activity, but clearly they enhance site II binding.

The relative efficiency of binding of *E. coli* T antigens to different templates also depended to some extent on the

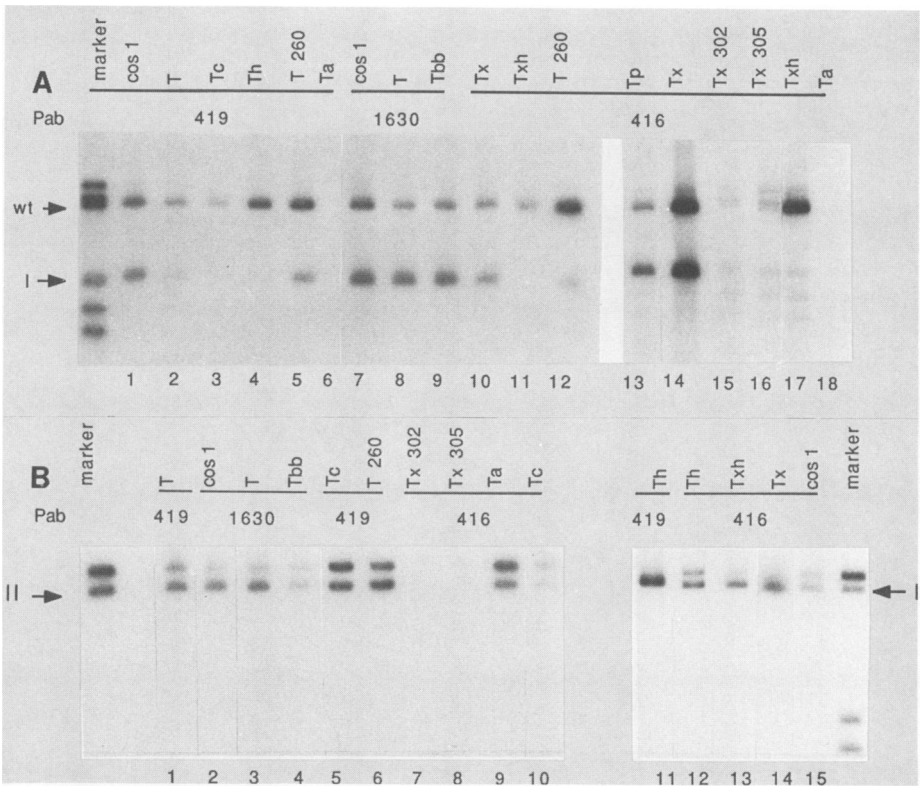

FIG. 5. Specific binding of T antigen expressed in *E. coli* to SV40 origin DNA. T antigen was immunoprecipitated from the indicated *E. coli* extracts with 5 μg of each of PAbs 419, 416, and 1630, as indicated. COS 1 extracts served as a control. A total of 250 ng of end-labeled plasmid DNA was added to each binding reaction, either (A) an equimolar mixture of site I (pONwt) and wild-type (wt) origin DNAs (sites I and II) or (B) p1097 (site II) DNA. The fragments containing each site are indicated by arrows. Note that the nonspecific band above the site II band in panel B is a doublet. Marker, 4% of the input DNA used in the binding reactions.

antibody chosen for immunopurification. Full-length *E. coli* T antigen and T260 bound site I with different relative efficiencies compared with the wild-type origin when different antibodies were used to immunoprecipitate the proteins (Fig. 5A, compare lanes 2 and 8 and lanes 5 and 12). The DNA-binding activity of COS 1 T antigen on site I and wild-type origin DNA, on the other hand, did not vary with the antibody used (Fig. 5A, lanes 1 and 7). The level of specific site II binding was also influenced in some cases by the antibody used for immunoprecipitation, for example COS 1 (Fig. 5B, lanes 2 and 15), T (lanes 1 and 3), and Th (lanes 11 and 12). These results suggest that the T antigen may exist in different conformations which are differentially recognized by the antibodies and have different DNA-binding activities. Alternatively, the antibodies may affect the conformation of the bound protein, favoring binding to site I or site II.

Purification of T-antigen peptides from *E. coli* extracts. Immunoaffinity chromatography on a Pab 419 antibody column (43) was ideally suited to purify full-length T-antigen gene products from the primary proteolytic degradation products, since these lacked the Pab 108 and 419 epitopes at the N terminus (Fig. 6A and not shown). Purification of the Th peptide is shown as an example. Samples from each stage of the purification procedure were analyzed by polyacrylamide gel electrophoresis (PAGE) and Coomassie staining (Fig. 6B) or Western blotting (Fig. 6C). The degradation products detected in the crude extract and early stages of purification by Pab 416 (Fig. 6A and C, lanes 1 to 4) were efficiently removed by immunoaffinity chromatography (Fig. 6B and C, lanes 6 to 9). At this stage, the main Th peptide represented greater than 90% of the purified protein (Fig. 6B and C, lanes 6 to 8). Further purification on an FPLC Mono Q column resulted in an apparently homogeneous preparation of Th (Fig. 6C, lane 11, and data not shown). The yield of purified Th at this stage from a 100-ml culture was about 40 μg. The purified material retained specific DNA-binding activity for SV40 origin DNA (Fig. 6D, lane 7).

DISCUSSION

By introducing new restriction sites at the intron boundaries of the large-T-antigen gene of SV40, it was possible to clone full-length large T antigen plus shorter regions into the procaryotic expression vector pUC9. Such a strategy, which could be generally applied to any intron-containing gene, offers a simple, error-free alternative to reverse transcription. Moreover, the new enzyme cleavage sites made it possible to manipulate the second exon of T antigen independently of the first exon sequences. With these and other new restriction sites, full-length large T antigen and several mutant proteins were inducibly expressed in *E. coli* as fusion polypeptides with the predicted sizes and epitopes detected by a panel of antibodies.

Initial functional analysis of T antigen made in *E. coli* indicates that the full-length protein binds specifically to SV40 origin DNA with an activity similar to that of the monkey cell protein. Since T antigen produced in *E. coli* is unlikely to be subject to the specific posttranslational modifications observed in the monkey cell protein (reviewed in

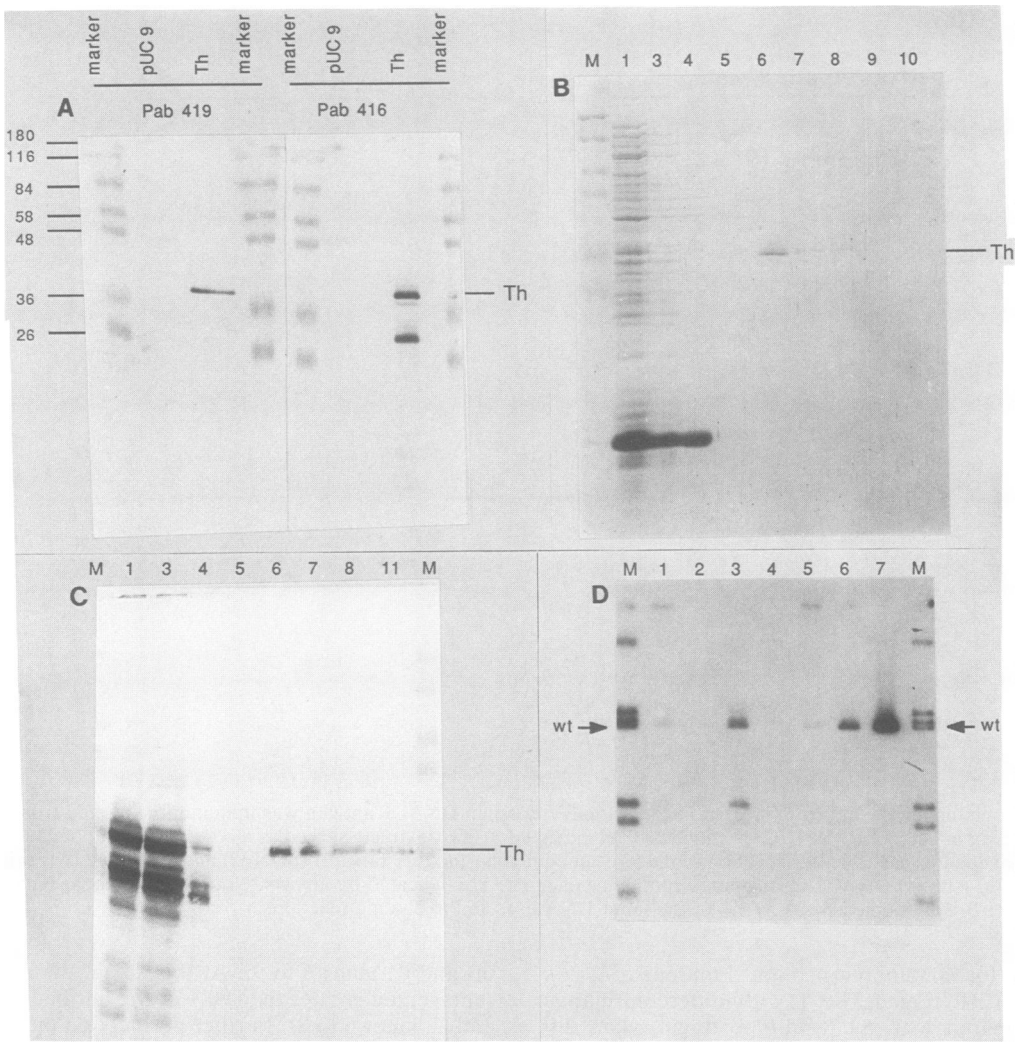

FIG. 6. Purification and DNA binding of Th antigen. (A) Crude extract from *E. coli* carrying Th or pUC9 was analyzed by immunoblotting with Pab 419 or 416 as indicated. Lanes marker or M, Prestained marker proteins (shown in kilodaltons). Equivalent samples of Th from each purification stage were analyzed by SDS-PAGE and Coomassie staining (B) or immunoblot with Pab 416 as the first antibody (C). Lane 1, Crude lysate; lane 2, flowthrough from Q-Sepharose FF; lane 3, eluate from Q-Sepharose FF; lane 4, flowthrough from Pab 419 affinity column; lane 5, wash from affinity column; lanes 6 to 9, eluate from Pab 419 column (fractions 3, 10, 20, and 30, respectively); lane 10, affinity column elution buffer; lane 11, eluate from FPLC Mono Q. (D) Origin-specific DNA binding of purified Th was assayed with end-labeled *Hin*dIII-digested pSVwt DNA, as in Fig. 5A. Lane 1, Supernatant of crude lysate; lanes 2 to 4, as in panel A; lanes 5 and 6, eluate from Pab 419 column, fractions 2 and 5, respectively. Lane 8, eluate from FPLC Mono Q. Lanes M, 5% of the input DNA used in the binding reaction. wt, Wild type.

reference 6), our results suggest that these modifications are not essential for origin DNA-binding activity. However, differences between the bacterial and monkey proteins in binding activity on isolated site I and II DNA may well derive from the lack of modification of the bacterial T antigen.

Truncated T-antigen mutants were analyzed to define a minimal peptide sequence required for specific SV40 origin DNA binding. The smallest specific DNA-binding peptide of T antigen was previously found to be a proteolytic fragment comprising residues 131 to 371 (44). The *E. coli* T antigen constructs that bore this sequence bound specifically to origin DNA, confirming that the origin-binding domain is localized in this region. Larger deletions of T-antigen coding sequences narrowed the origin DNA-binding domain to a region within residues 131 to 259. Furthermore, one peptide, Th, purified to greater than 90% homogeneity, retained specific origin DNA-binding activity, confirming the potential of this approach to produce large amounts of biologically active peptides. DNA binding of similar peptides expressed in *E. coli* has also been shown independently by other groups (26a, 48). These results provide positive functional evidence to support the deductions made from studies of mutant proteins defective in origin DNA-binding activity (29, 32, 49).

Clearly, however, the presence or absence of other regions of the protein affects its stability, binding activity, and preference for binding to site I or II. More detailed binding studies with purified peptides will be required for quantitative comparisons of site I and site II binding. The lack of modification of the truncated peptides produced in *E. coli* may also affect their relative binding activities on sites I and II. Recent studies of the DNA-binding properties of enzymatically dephosphorylated T antigen and mutant T antigens

carrying conservative amino acid substitutions in phosphorylated residues point to a role for phosphorylation in differential recognition of sites I and II (26, 39, 45).

The present data do not allow a more precise localization of the residues within the region 131 to 259 which are involved in recognizing the consensus pentanucleotide binding signals. A mutational analysis of site I demonstrated the tripartite nature of the binding site, in which the sequence between the two pentanucleotides is thought to determine a bent DNA conformation required for optimal binding of two T-antigen monomers (37, 38). These data were interpreted to suggest that not only protein-DNA interactions but also protein-protein interactions between the T-antigen subunits are required for optimal binding. If this interpretation is correct, then the same 129-amino-acid region of T antigen is likely to be responsible for both the protein-DNA and the protein-protein interactions. The mutant Ta (residues 1 to 239) did not have sufficient origin DNA-binding activity to be detected by the rather stringent binding assay used. However, it is not yet clear whether DNA-protein interactions, protein-protein interactions, or the overall conformation of the peptide is disrupted in this mutant. More extensive mutational analysis of the minimal origin-binding domain, perhaps in conjunction with detailed binding studies on wild-type and mutant minimal binding sites (37, 38), may reveal more clearly the interactions involved in T-antigen DNA binding.

While the specific origin-binding domain is clearly localized within residues 131 to 259, the downstream region between residues 260 and 371, previously reported to contribute to nonspecific cellular DNA binding (31), contains a sequence homologous to one zinc finger motif. This sequence, Cys-X_2-Cys-X_{11}-His-X_2-His, is precisely conserved in the same relative location in six papovavirus T antigens (J. M. Pipas, personal communication). Replacement of either of the two cysteine residues, Cys-302 or Cys-305, by serine abolished specific origin DNA-binding activity. However, deletion of the DNA region encoding the mutant finger sequences restored origin DNA-binding activity.

These data suggest that the finger structures postulated for eucaryotic DNA-binding proteins (2, 9) need not always be directly involved in specific DNA-protein interaction. Alternatively, we suggest that these cysteine residues in T antigen may be structurally important, for example, for intramolecular disulfide bonds (44), which could play a role in folding or stabilizing the protomer (18). In agreement with this interpretation, the *tsA*30 mutation was mapped at residue 300 adjacent to the finger motif (6). However, we cannot exclude the possibility that SV40 T antigen contains a second DNA-binding domain involved in non-sequence-specific binding or even specific binding to as yet undefined sequences in cellular DNA.

In summary, cloning and expression of SV40 large T antigen in *E. coli* has facilitated the positive identification of a small region responsible for specific binding of the protein to site I and site II of SV40 origin DNA. We anticipate that this expression system can be exploited for genetic analysis of T-antigen structure and function, as a convenient source of biochemically functional peptides in quantities sufficient for structural studies, and perhaps as a model system for the development of new cloning strategies for genes encoding eucaryotic DNA-binding proteins.

ACKNOWLEDGMENTS

We thank Ronald Mertz and Dorit Weigand for oligonucleotide synthesis, Robert Garcea for *E. coli* SG935IQ, Silke Dehde for valuable technical assistance, and Maria Ihmsen for excellent secretarial assistance. We are grateful to Robert Runzler, Jens Schneider, Josee Espen, and Karl-Heinz Westphal for helpful advice.

The financial support of the Bundesministerium für Forschung und Technologie (Genzentrum), Deutsche Forschungsgemeinschaft (Fa 138-1/2 and 3/1), and Fonds der Chemischen Industrie is gratefully acknowledged.

LITERATURE CITED

1. **Ball, R. K., B. Siegl, S. Quellhorst, G. Brandner, and D. G. Braun.** 1984. Monoclonal antibodies against simian virus 40 large T antigen: epitope mapping, papova virus cross-reaction and cell surface staining. EMBO J. **3**:1485–1491.
2. **Berg, J.** 1986. Potential metal-binding domains in nucleic acid binding proteins. Science **232**:485–486.
3. **Brady, J., and G. Khoury.** 1985. *trans*-Activation of the simian virus 40 late transcription unit by T antigen. Mol. Cell. Biol. **5**:1391–1399.
4. **Clark, R., K. Peden, J. Pipas, D. Nathans, and R. Tjian.** 1983. Biochemical activities of T-antigen proteins encoded by simian virus 40 *A* gene deletion mutants. Mol. Cell. Biol. **3**:220–228.
5. **DeLucia, A. L., B. A. Lewton, R. Tjian, and P. Tegtmeyer.** 1983. Topography of simian virus 40 protein-DNA complexes: arrangement of pentanucleotide interaction sites at the origin of replication. J. Virol. **46**:143–150.
6. **DePamphilis, M. L., and M. K. Bradley.** 1986. Replication of SV40 and polyoma virus chromosomes, p. 99–246. *In* N. P. Salzman (ed.), The papovaviridae, vol. 1. Plenum Publishing Co., New York.
7. **DiMaio, D., and D. Nathans.** 1982. Regulatory mutants of simian virus 40. Effect of mutations at a T-antigen binding site on DNA replication and expression of viral genes. J. Mol. Biol. **156**:531–548.
8. **Dretzen, G., M. Bellard, P. Sassone-Corsi, and P. Chambon.** 1981. A reliable method for the recovery of DNA fragments from agarose and acrylamide gels. Anal. Biochem. **112**:295–298.
9. **Evans, R. M., and S. M. Hollenberg.** 1988. Zinc fingers: guilt by association. Cell **52**:1–3.
10. **Fanning, E., B. Nowak, and C. Burger.** 1981. Detection and characterization of multiple forms of simian virus 40 large T antigen. J. Virol. **37**:92–102.
11. **Fanning, E., K.-H. Westphal, D. Brauer, and D. Cörlin.** 1982. Subclasses of simian virus 40 large T antigen: differential binding of two subclasses of T antigen from productively infected cells to viral and cellular DNA. EMBO J. **1**:1023–1028.
12. **Goff, S. A., L. P. Casson, and A. L. Goldberg.** 1984. Heat shock regulatory gene *htp*R influences rates of protein degradation and expression of the *ion* gene in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **81**:6647–6651.
13. **Gurney, E. G., S. Tamowsky, and W. Deppert.** 1986. Antigenic binding sites of monoclonal antibodies specific for simian virus 40 large T. J. Virol. **57**:1168–1172.
14. **Hansen, U., D. Tenen, D. M. Livingston, and P. A. Sharp.** 1981. T antigen repression of SV40 early transcription from two promoters. Cell **27**:603–612.
15. **Harlow, E., L. V. Crawford, D. C. Pimand, and N. M. Williamson.** 1981. Monoclonal antibodies specific for simian virus 40 tumor antigens. J. Virol. **39**:861–869.
16. **Jones, K. A., and R. Tjian.** 1984. Essential contact residues within SV40 large T antigen binding sites I and II identified by alkylation-interference. Cell **36**:155–162.
17. **Keller, J. M., and J. C. Alwine.** 1985. Analysis of an activatable promoter: sequences in the simian virus 40 late promoter required for T-antigen-mediated *trans*-activation. Mol. Cell. Biol. **5**:1859–1869.
18. **King, J.** 1986. Genetic analysis of protein folding pathways. Bio/Technology **4**:297–303.
19. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
20. **Levinson, A., D. Silver, and B. Seed.** 1984. Minimal size plasmids containing an M13 origin for production of single-strand transducing particles. J. Mol. Appl. Genet. **2**:507–517.

21. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
22. **Manos, M., and Y. Gluzman.** 1985. Genetic and biochemical analysis of transformation-competent, replication-defective simian virus 40 large-T-antigen mutants. J. Virol. **53**:120–127.
23. **Margolskee, R., and D. Nathans.** 1984. Simian virus 40 mutant T antigens with relaxed specificity for the nucleotide sequence at the viral origin of DNA replication. J. Virol. **49**:386–393.
24. **Mastrangelo, I. A., P. V. C. Hough, V. G. Wilson, J. S. Wall, J. F. Hainfeld, and P. Tegtmeyer.** 1985. Monomers through trimers of large tumor antigen bind in region II of simian virus 40 origin of replication DNA as stable structures in solution. Proc. Natl. Acad. Sci. USA **82**:3626–3630.
25. **Miller, J., A. D. McLachlan, and A. Klug.** 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. EMBO J. **4**:1609–1614.
26. **Mohr, I. J., B. Stillman, and Y. Gluzman.** 1987. Regulation of SV40 DNA replication by phosphorylation of T antigen. EMBO J. **6**:153–160.

26a. **Mole, S. E., J. V. Gannon, M. J. Ford, and D. P. Lane.** 1987. Structure and function of large T antigen. Philos. Trans. R. Soc. London B Biol. Sci. **317**:455–469.

27. **Morrison, B., M. Kress, G. Khoury, and G. Jay.** 1983. Simian virus 40 tumor antigen: isolation of the origin-specific DNA-binding domain. J. Virol. **47**:106–114.
28. **Myers, R., and R. Tjian.** 1980. Construction and analysis of simian virus 40 origins defective in tumor antigen binding and DNA replication. Proc. Natl. Acad. Sci. USA **77**:6491–6495.
29. **Paucha, E., D. Kalderon, R. Harvey, and A. E. Smith.** 1986. Simian virus 40 origin DNA-binding domain on large T antigen. J. Virol. **57**:50–64.
30. **Pipas, J. M., K. W. C. Peden, and D. Nathans.** 1983. Mutational analysis of simian virus 40 T antigen: isolation and characterization of mutants with deletions in the T-antigen gene. Mol. Cell. Biol. **3**:204–213.
31. **Prives, C., B. Barnet, A. Scheller, G. Khoury, and G. Jay.** 1982. Discrete regions of simian virus 40 large T antigen are required for nonspecific and viral origin-specific DNA binding. J. Virol. **43**:73–82.
32. **Prives, C., L. Covey, A. Scheller, and Y. Gluzman.** 1983. DNA-binding properties of simian virus 40 T-antigen mutants defective in viral DNA replication. Mol. Cell. Biol. **3**:1958–1966.
33. **Ptashne, M.** 1986. A genetic switch. Blackwell Scientific Publications, Cambridge, Mass.
34. **Rigby, P., and D. Lane.** 1983. The structure and function of SV40 large T antigen. Adv. Viral Oncol. **3**:31–57.
35. **Rio, D., A. Robbins, R. Myers, and R. Tjian.** 1980. Regulation of simian virus 40 early transcription *in vitro* by a purified tumor antigen. Proc. Natl. Acad. Sci. USA **77**:5706–5710.
36. **Runzler, R., S. Thompson, and E. Fanning.** 1987. Oligomerization and origin DNA-binding activity of simian virus 40 large T antigen. J. Virol. **61**:2076–2083.
37. **Ryder, K., E. Vakalopoulou, R. Mertz, I. Mastrangelo, P. Hough, P. Tegtmeyer, and E. Fanning.** 1985. Seventeen base pairs of region I encode a novel tripartite binding signal for SV40 T antigen. Cell **42**:539–548.
38. **Ryder, K., S. Silver, A. L. DeLucia, E. Fanning, and P. Tegtmeyer.** 1986. An altered DNA conformation in origin region I is a determinant for the binding of SV40 large T antigen. Cell **44**:719–725.
39. **Schneider, J., and E. Fanning.** 1988. Mutations in the phosphorylation sites of simian virus 40 (SV40) T antigen alter its origin DNA-binding specificity for site I or II and affect SV40 DNA replication activity. J. Virol. **62**:1598–1605.
40. **Schutzbank, T., R. Robinson, M. Oren, and A. J. Levine.** 1982. SV40 large tumor antigen can regulate some cellular transcripts in a positive fashion. Cell **30**:481–490.
41. **Scott, M. R. D., K.-H. Westphal, and P. W. J. Rigby.** 1983. Activation of mouse genes in transformed cells. Cell **34**:557–567.
42. **Shortle, D., R. Margolskee, and D. Nathans.** 1979. Mutational analysis of the simian virus 40 replicon: pseudorevertants of mutants with a defective replication origin. Proc. Natl. Acad. Sci. USA **76**:6128–6131.
43. **Simanis, V., and D. P. Lane.** 1985. An immunoaffinity purification procedure for SV40 large T antigen. Virology **144**:88–100.
44. **Simmons, D. T.** 1986. DNA-binding region of the simian virus 40 tumor antigen. J. Virol. **57**:776–785.
45. **Simmons, D. T., W. Chou, and K. Rodgers.** 1986. Phosphorylation downregulates the DNA-binding activity of simian virus 40 T antigen. J. Virol. **60**:888–894.
46. **Soprano, K. J., N. Galanti, G. J. Jonak, S. McKercher, J. M. Pipas, K. W. C. Peden, and R. Baserga.** 1983. Mutational analysis of simian virus 40 T antigen: stimulation of cellular DNA synthesis and activation of rRNA genes by mutants with deletions in the T-antigen gene. Mol. Cell. Biol. **3**:214–219.
47. **Stillman, B., R. D. Gerard, R. A. Guggenheimer, and Y. Gluzman.** 1985. T antigen and template requirements for SV40 DNA replication *in vitro*. EMBO J. **4**:2933–2939.
48. **Strauss, M., P. Argani, I. J. Mohr, and Y. Gluzman.** 1987. Studies on the origin-specific DNA-binding domain of simian virus 40 large T antigen. J. Virol. **61**:3326–3330.
49. **Stringer, J. R.** 1982. Mutant of simian virus 40 large T antigen that is defective for viral DNA synthesis but competent for transformation of cultured rat cells. J. Virol. **42**:854–864.
50. **Tegtmeyer, P., B. A. Lewton, A. L. DeLucia, V. G. Wilson, and K. Ryder.** 1983. Topography of simian virus 40 A protein-DNA complexes: arrangement of protein bound to the origin of replication. J. Virol. **46**:151–161.
51. **Tjian, R.** 1978. The binding site of SV40 DNA for a T antigen-related protein. Cell **13**:165–179.
52. **Towbin, H., T. Staehelin, and J. Gordon.** 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA **76**:4350–4354.
53. **Vieira, J., and J. Messing.** 1982. The pUC plasmids: an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:259–268.
54. **Vogt, B., E. Vakalopoulou, and E. Fanning.** 1986. Allosteric control of simian virus 40 T antigen binding to viral-origin DNA. J. Virol. **58**:765–772.
55. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.
56. **Zoller, M. J., and M. Smith.** 1982. Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any fragment of DNA. Nucleic Acids Res. **10**:6487–6500.