JOURNAL OF VIROLOGY, May 1990, p. 1973–1983
0022-538X/90/051973-11$02.00/0

Vol. 64, No. 5

Four Major Sequence Elements of Simian Virus 40 Large T Antigen Coordinate Its Specific and Nonspecific DNA Binding

DANIEL T. SIMMONS,[1,2]* GERHARD LOEBER,[2]† AND PETER TEGTMEYER[2]

School of Life and Health Sciences, University of Delaware, Newark, Delaware 19716,[1] and Department of Microbiology, State University of New York, Stony Brook, New York 11794[2]*

Received 17 October 1989/Accepted 23 January 1990

By mutational analysis, we have identified a motif critical to the proper recognition and binding of simian virus 40 large tumor antigen (T antigen) to virus DNA sequences at the origin of DNA replication. This motif is tripartite and consists of two elements (termed A1 and B2) that are necessary for sequence-specific binding of the origin and a central element (B1) which is required for nonspecific DNA-binding activity. Certain amino acids in elements A1 (residues 152 to 155) and B2 (203 to 207) may make direct contact with the GAGGC pentanucleotide sequences in binding sites I and II on the DNA. Alternatively, these two elements could determine the proper structure of the DNA-binding domain, although for a number of reasons we favor the first possibility. In contrast, element B1 (183 to 187) is most likely important for recognizing a general structural feature of DNA. Elements A1 and B2 are nearly identical in all known papovavirus T antigens, whereas B1 is identical only in the closely related papovaviruses simian virus 40, BK virus, and JC virus. In addition to these three elements, a fourth (B3; residues 215 to 219) is necessary for the binding of T antigen to site II but not to site I. We propose that additional contact sites on T antigen are involved in the interaction with site II to initiate the replication of the viral DNA.

The simian virus 40 (SV40) large tumor antigen (T antigen) is a 708-amino-acid phosphoprotein that is involved in the initiation and elongation stages of SV40 DNA synthesis in infected monkey cells. Its role in DNA replication is beginning to be understood. It appears that T antigen first binds to the pentanucleotide sequences GAGGC in sites I and II close to the replication origin (10, 12, 20, 47, 48) and melts the DNA in the proximity of site II (4). After DNA synthesis has begun, T antigen may function as a helicase to unwind the parental DNA strands (8, 14, 43, 49).

The mechanism by which papovavirus T antigens bind to the sequences at the replication origin is not known. Over the last few years, the domain responsible for DNA binding has been identified by genetic (7, 32, 34) and biochemical (39) approaches. However, little or no similarity exists between this domain and that of well-studied DNA-binding proteins, so no obvious model presently exists to explain the binding to DNA.

There is considerable evidence which implicates the region between residues 140 and 260 in origin binding (1, 40, 46). A Zn^{2+} finger motif between residues 301 and 321 appears to contribute to the stability of the DNA-protein complex (1, 40), although it is not required for DNA binding per se.

Several groups have generated mutants of T antigen with altered binding activity for the origin. With one notable exception (32), however, their approach has been to study only one or a few mutants within this region, and consequently there is very little information about the overall organization of the DNA-binding domain. We have started a systematic and thorough mutagenesis of this domain in order to identify the sequences that are important for *ori*-binding activity. The aim here is twofold. One goal is to identify amino acid residues that make contact with sequences at the origin, and the second is to find the residues that are crucial for the proper structure (and therefore function) of this domain. With this information, a model of DNA binding may emerge.

Our survey has revealed that the DNA-binding domain has a complex organization. By testing the effects of single-site mutations on virus replication and by performing a number of DNA-binding assays on replication-negative T-antigen mutants, we have identified several important sequence elements within this domain. This information has been used to develop a basic model that describes the interaction of the domain with origin and nonorigin DNA.

MATERIALS AND METHODS

Plasmids. pBS-SV40 contains the entire SV40 genome inserted into the *Bam*HI site of Bluescript (Stratagene). pSKAT is a plasmid derived in our lab from pIA#4 (a gift from Y. Gluzman). It contains the SV40 T-antigen gene inserted between adenovirus type 5 map units 0 to 1.4 and the major late promoter of adenovirus type 2. Additional adenovirus sequences (map units 9 to 15.5) are present to allow for recombination with homologous sequences in a large adenovirus fragment (map units 4 to 100). pSVO+ contains the wild-type SV40 origin consisting of sites I and II; pOS1 contains site I only, and pSVOdl3 contains site II only (44).

Mutagenesis protocol. Mutations were generated in pBS-SV40 or pSKAT by annealing oligonucleotides with a single mismatch to a uridine-containing single-stranded DNA template as previously described (23). The oligonucleotide was extended with T4 DNA polymerase, and the resulting double-stranded DNA was used to transform *Escherichia coli* BMH 71-18 (International Biotechnologies, Inc.). Single-stranded DNA was sequenced by the dideoxy procedure (37). In some cases the mutant T antigen gene was recloned from pBS-SV40 to pSKAT by standard recombinant DNA procedures.

Virus replication assays. pBS-SV40 harboring a mutation

* Corresponding author.
† Present address: Ernst Boehringer Institute, Bender-AG, Dr. Boehringergasse 5-11, A-1121 Vienna, Austria.

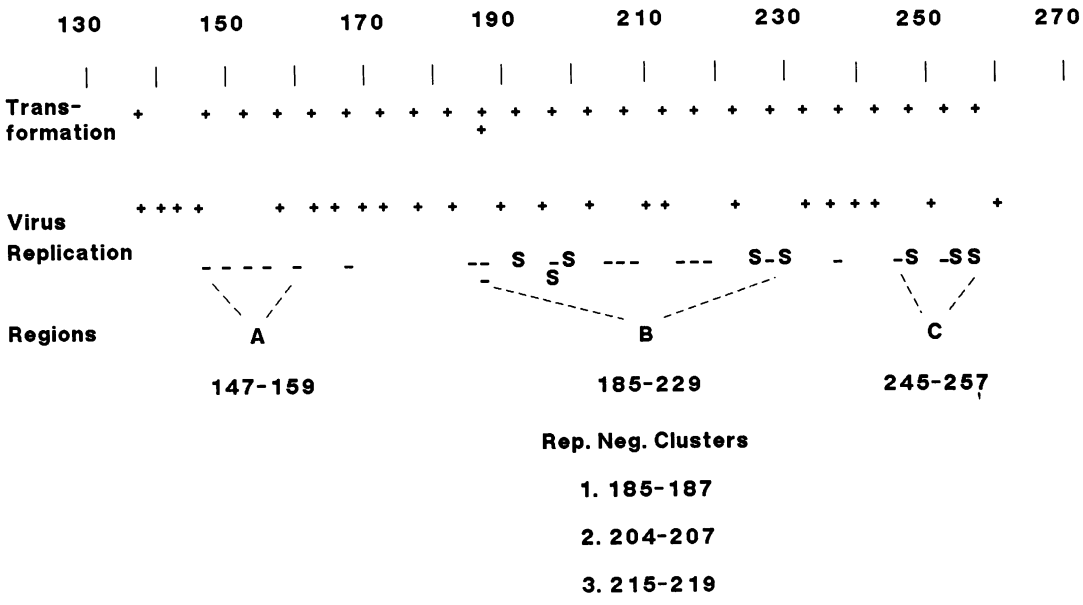

FIG. 1. Plaque formation and transformation assays of T-antigen mutants. Single-point substitution mutations in the DNA-binding domain were tested for their effects on virus multiplication in monkey cells. Symbols: +, mutations that resulted in replication like that of wild-type SV40; S, mutations that gave rise to small plaques; −, mutations that resulted in no replication. Mutations that had an effect on virus replication are shown on a separate line from those that had no effect. Mutants with altered replication properties were arranged into three groups (A, B, and C), with the central group (B) containing three clusters of replication-negative (rep. neg.) mutants. Some of the mutants were also tested for the ability to transform primary mouse embryo cells (transformation +). In some cases, we tested two mutants with different amino acid substitutions at the same site.

in the gene for T antigen was cleaved with *Bam*HI to release the mutant genomic DNA. The DNA was ligated at low DNA concentrations to favor the formation of circular DNA and then transfected into monkey cells (CV-1 or BSC-1) by the DEAE-dextran procedure (27) as previously described (23). Plaques were counted 10 to 30 days posttransfection, depending upon plaque size. Plates which did not have any plaques were incubated for a minimum of 30 days to make sure that small plaques did not appear.

Transformation assays. Ligated circular SV40 DNAs containing point mutations in the DNA-binding domain were tested for their transforming activities on primary mouse embryo cells as described previously (23).

Adenovirus recombinants. Adenovirus-SV40 recombinants containing the gene for T antigen (mutant or wild type) were generated by cotransfection of *Kpn*I-linearized pSKAT and *Xba*I fragment A of adenovirus type 5 *dl*309 (21) (map units 4 to 100) in adenovirus-transformed 293 cells as described previously (3, 25, 42, 45). Transfections were carried out by using a $CaPO_4$ precipitation technique (6). Plaques were picked, and the virus was grown out and screened for T-antigen expression by immunofluorescence. Virus stocks giving rise to 50 to 100% T-antigen-positive cells were usually used for T-antigen production. Recombinant viruses were obtained for all replication-negative mutations but one (185ST).

Preparation of mutant T antigen. 293 cells were infected with recombinant adenoviruses at a multiplicity of about 10 PFU per cell. At 20 h, the cells were lysed and T antigen was recovered from the lysate by immunoprecipitation (41) with PAb416 monoclonal antibody. In the case of the 185ST mutation, sufficient T antigen for biochemical analysis was obtained by transfection of the linearized pSKAT into 293 cells by using the same protocol as that for generating recombinant adenovirus. The binding of this mutant T antigen to DNA was compared with that of wild-type T antigen prepared in the same way.

DNA-binding assays. In most experiments, DNA-binding activity was measured as previously described (41), using a modification of the assay of McKay (28). All binding reactions were performed at T-antigen excess. DNA-binding activity under replication conditions was measured by the method described by Deb and Tegtmeyer (11). When present, competitor DNA was added from the beginning of the reaction.

RESULTS

Generation of mutations in the DNA-binding domain. We generated single-point mutants at regular intervals in the DNA-binding domain of T antigen. Mutations were initially made at every 5 amino acid residues. Additional mutations were made in a second round of mutagenesis near sites which had an effect on virus replication. The substitutions were chosen so as to make conservative changes. Wherever possible, the same substitution was made for one amino acid; however, different substitutions were sometimes made when different codons were involved because, for efficiency, we wanted to limit the mutation in the DNA to a single base pair. Our aim in making the "softest" possible amino acid substitutions was to keep structural alterations in the protein to a minimum. This, we reasoned, would optimize our chances of identifying residues that make contact with the GAGGC pentanucleotides at the origin.

Mutations were generated in a plasmid containing the entire SV40 genome (pBS-SV40) (23). The mutagenesis protocol was similar to the one devised by Kunkel (22) and is described in detail elsewhere (23). Mutant SV40 DNA was excised from the plasmid and then tested for infectivity in monkey cells. Virus replication results for 51 mutants that

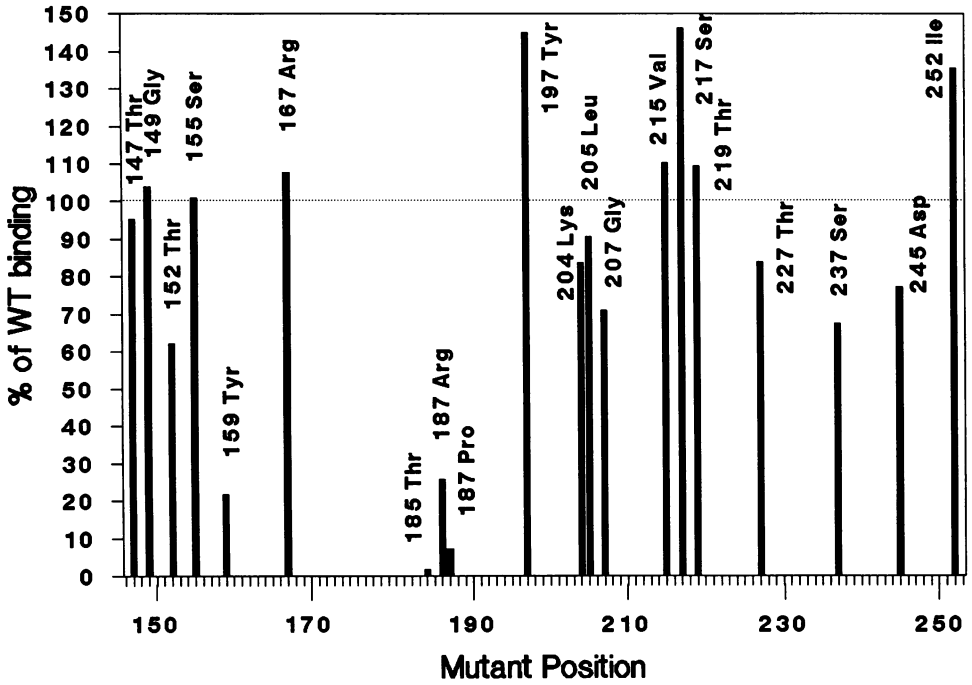

FIG. 2. Binding of mutant T-antigen proteins to plasmid DNA. Mutant T antigens were immunoprecipitated with PAb416 and tested for binding to a small (400-base-pair) labeled plasmid fragment (*Taq*I fragment D of pSVO+). Results are shown as a percentage of wild-type T-antigen-binding activity under identical conditions (competitor calf thymus DNA was not added); 100% represents about 10,000 cpm of labeled DNA bound. Results are also shown in Table 1.

were made in the DNA-binding domain are shown in Fig. 1. About 45% of the mutants replicated like wild-type virus, 15% replicated at less than wild-type rates as judged by the appearance of small plaques (S), and the remaining 40% did not replicate at all even at high input DNA concentrations. Figure 1 also shows that all of the mutants that were tested transformed primary mouse embryo cells. There was no significant qualitative or quantitative difference in the transforming properties of these mutants relative to that of wild-type virus, indicating that gross structural alterations did not occur in the mutant T antigens.

We found it useful to group the mutants that showed an effect on virus replication into three classes corresponding to regions in the DNA-binding domain, which we term A, B, and C. (Fig. 1). These regions mapped from residue 147 to 159, 185 to 229, and 245 to 257, respectively. Only one replication-negative mutation (at residue 167) was localized between regions A and B, and another (at residue 237) was between regions B and C. The residue 237 mutant was actually unimpaired in DNA binding (see below), which makes the division between regions B and C somewhat clearer. The large central region contained several clusters of replication-negative mutants mapping at residues 185 to 187, 203 to 207, and 215 to 219.

Generation of mutant proteins. The mutants that did not replicate in monkey cells are likely to be the ones with altered domain function. T antigens harboring these mutations were prepared for biochemical analysis. For this purpose, corresponding-base-pair mutations were made in an adenovirus-SV40 hybrid virus which was used as an expression vector. Mutant T-antigen proteins were prepared by immunoprecipitation (41) from infected 293 cells by using PAb416 (18) monoclonal antibody. The amounts of T antigen in each sample were estimated by Coomassie blue staining of acrylamide gels followed by densitometry. Equal amounts of each T antigen were used in binding assays.

DNA-binding assays of mutant T antigens. We have previously adopted a rapid and quantitative assay for DNA binding using T antigen bound to *Staphylococcus aureus* (41). This assay was used for comparing the DNA-binding activities of mutant proteins with that of wild-type T antigen. Several different labeled substrate DNAs and two different conditions were used. First, to measure binding of mutant proteins to any double-stranded DNA, we used a small (about 400-base-pair) fragment containing plasmid sequences. Binding with this DNA was performed without any added competitor DNA (Fig. 2 and Table 1). All but four mutant proteins demonstrated binding activity close to that of wild type. The four mutations mapped at residues 159, 185, and 187 (His to Arg and His to Pro), and the protein had activity which was 25% or less of that of the wild type (Table 1 contains a complete list of amino acid substitutions in the replication-negative mutants). The four mutant T antigens are classified as poor nonspecific DNA binders.

We next tested the activities of the mutant T antigens for binding to a labeled DNA fragment containing the wild-type SV40 DNA origin region (sites I and II) (Fig. 3A and B and Table 1). The assay was performed in the absence or presence of unlabeled competitor (1,000-fold excess) calf thymus DNA. In the first case, both *ori*- and nonspecific-binding activities were measured, and in the second, only *ori*-specific binding was measured (34, 41, and unpublished results) because the calf thymus DNA competes with nonspecific binding. When the binding of each mutant protein to the ori fragment was compared in the absence and presence of competitor DNA, two major differences appeared (Fig. 3A and B and Table 1). The activities of the proteins with mutations at residues 155 and 204 dropped significantly in

TABLE 1. DNA-binding results of replication-negative T-antigen mutants

Residue no. and type in WT[a]	Mutation	Mutant name	% of WT T antigen binding to[b]:							Effect[e]
			Plasmid (no comp)	WT DNA (no comp)	WT DNA	WT DNA[c]	Site I	Site II	Site II-rep[d]	
147 Ser	Thr	147ST	95	57	44	42	19	9	20	I, II
149 Ala	Gly	149AG	104	113	103	102	84	69	91	None
152 Ser	Thr	152ST	62	67	49	48	39	13	18	I, II
155 Thr	Ser	155TS	101	48	6	6	2	1	11	I, II
159 Phe	Tyr	159FY	22	21	21	20	23	4	18	All
167 Lys	Arg	167KR	108	133	99	77	86	23	35	II
185 Ser	Thr	185ST	0	1	1	1	0	0	0	All
187 His	Arg	187HR	26	4	2	0	0	8	17	All
187 His	Pro	187HP	7	2	1	0	0	2	16	All
197 Phe	Tyr	197FY	145	146	124	121	83	86	104	None
204 Arg	Lys	204RK	84	39	1	4	1	2	13	I, II
205 Val	Leu	205VL	81	92	88	69	63	18	36	II
207 Ala	Gly	207AG	71	72	37	34	18	11	22	I, II
215 Leu	Val	215LV	110	131	102	84	92	29	39	II
217 Thr	Ser	217TS	146	139	100	94	84	40	50	II
219 Ser	Thr	219ST	109	119	99	86	81	22	35	II
227 Asn	Thr	227NT	84	68	47	42	21	10	19	I, II
237 Thr	Ser	237TS	67	115	115	113	108	89	104	None
245 Glu	Asp	245ED	77	99	87	76	82	31	46	II
252 Leu	Ile	252LI	135	124	76	71	60	48	51	I, II

[a] WT, Wild type.

[b] Mutant T antigens were isolated by immunoprecipitation from infected or (for Thr-185) transfected 293 cells. The proteins were then tested in DNA-binding reactions as described in Materials and Methods. Various substrate DNAs were used in the binding reactions as indicated. Calf thymus DNA was included (1,000-fold excess) as a competitor except where indicated (no comp).

[c] This assay with wild-type DNA was performed under the same conditions as the assays with site I and site II DNAs.

[d] Binding to site II DNA under replication conditions.

[e] Summary of effects on DNA binding to various substrate DNAs. I, II means that binding to both sites I and II were affected, and all means that binding to all sites was affected. The effects of the mutation at site 252 were small.

the presence of competitor DNA, demonstrating that these two sites (Thr-155 and Lys-204 in the wild-type protein) are important for origin-specific recognition or for the stability of the *ori* DNA-protein complex. The binding activity of the mutant with a change at residue 207 also dropped in the presence of competitor DNA, but the effect was not as dramatic. It should be noted that the activity of the protein with a mutation at residue 159 did not change from its original level of 20% (of that of wild type) (Fig. 2) when it was tested with *ori* DNA either with or without competitor DNA (Fig. 3 and Table 1).

Binding of mutant T antigens to DNA sites I and II. We next tested the binding activities of the mutant proteins to separated sites I and II. This is important since, under the conditions used, most of the binding to the origin is for site I sequences (17, 47, 48; unpublished results). The binding reactions were performed with DNA fragments containing deletions in sites II and I, respectively (44, 46). The binding activities of these mutant proteins to wild-type *ori* (Fig. 4A), to site I (Fig. 4B), and to site II (Fig. 4C) under identical conditions in the presence of competitor DNA are compared in Fig. 4 and Table 1. Several of the mutant proteins were placed in one group on the basis that they bound *ori* DNA at levels somewhat less than that of wild type and on the fact that they demonstrated a progressively lower activity as the substrate was changed from complete *ori* DNA to site I and then to site II. These are the five mutant proteins with changes at residues 147, 152, 207, 227, and 252. We can categorize these mutants as being affected in binding to sites I and II, although not all were affected equally. A second class of mutants demonstrated lowered binding to site II, mostly. The mutants in this group had changes at residues 167, 205, 215, 217, 219, and 245. Third, there were mutations (at residues 155, 159, 185, 187 [Arg and Pro], and 204) which resulted in low activity with all three substrates used in this experiment. This category includes the four poor nonspecific DNA binders (Fig. 2) as well as the two (at residues 155 and 204) which did not bind specifically to sequences at the origin (Fig. 3). In the fourth class, three mutations (at residues 149, 197, and 237) resulted in near-wild-type levels of activity with all three DNAs.

To demonstrate that the three mutants in the fourth class could bind to site II DNA under conditions that would permit the replication of the DNA in vitro, we repeated the binding under replication conditions (5, 11). Under these conditions, which include ATP and incubation at 37°C, binding to site II is enhanced compared with binding without ATP (5, 11). These three mutant proteins (with mutations at residues 149, 197, and 237, respectively) had wild-type levels of binding by this criterion (Fig. 5). All other mutant T antigens bound site II less well. The three mutant proteins with wild-type levels of DNA binding were deficient in helicase activity, which explains the replication defects of the mutant viruses (K. Wun-Kim and D. T. Simmons, unpublished data).

Table 1 summarizes the results of all DNA-binding assays. Also shown is a synopsis of the effects of each mutation on the binding to various substrate DNAs.

DISCUSSION

The effect of amino acid substitutions on DNA binding can be due to two different types of changes: a change in protein conformation or a change in a residue that is involved in function. It has been documented that changes in structure often affect the stability of proteins (31). Thus, for a protein whose structure is not known, it can be generally assumed that mutations which affect activity without having an effect

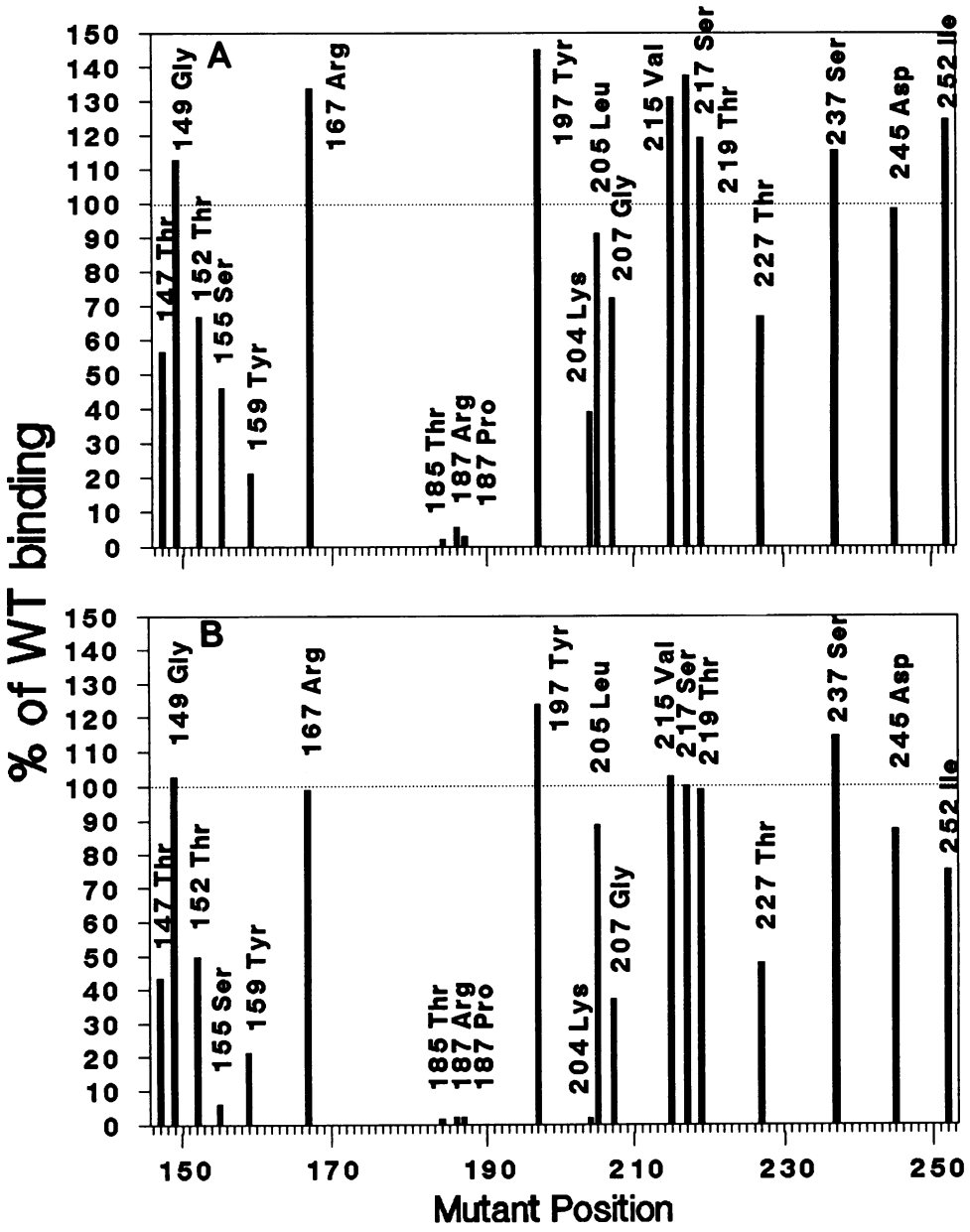

FIG. 3. Binding of mutant T antigens to wild-type SV40 origin. Immunoprecipitated mutant T antigens were tested in a binding assay with a labeled fragment containing the wild-type SV40 origin (*Taq*I fragment E of pSVO+). Reactions were performed in the absence (A) or presence (B) of 1,000-fold mass excess of unlabeled sheared calf thymus DNA as a competitor of non-sequence-specific binding.

on overall stability are in functional rather than structural sites (31). Although this is not a hard-and-fast rule, we have made use of it to develop a working model of T antigen-DNA interactions. Our mutant T antigens do not appear to have widely different stabilities, as judged by the facts that the mutants that were tested gave rise to an equal number of transformed cell foci in transformation assays (Fig. 1) and that about equal amounts of T antigen were generated by transfection in 293 cells (not shown). However, we do not want to exclude the possibility that, in some cases, changes in DNA-binding activity are due primarily to structural alterations.

The interpretation of our data also rests in part on the assumption that if several neighboring amino acids are found to be important for activity, the region itself is functionally involved. Our reasoning is that if the region is important only for determining the proper structure (i.e., the active site is elsewhere), a number of amino acid substitutions would be tolerated and, in our analysis, mutations resulting in phenotypic changes would not cluster within it. By making conservative amino acid changes, we have attempted to minimize signals due to conformational changes. The contention that a region serves a distinct function can be strengthened if all mutations in the region lead to similar phenotypes, i.e., the presence of one activity but not another. We have also considered the magnitude of the change in DNA binding to evaluate the possibility that any one amino acid is part of an active site.

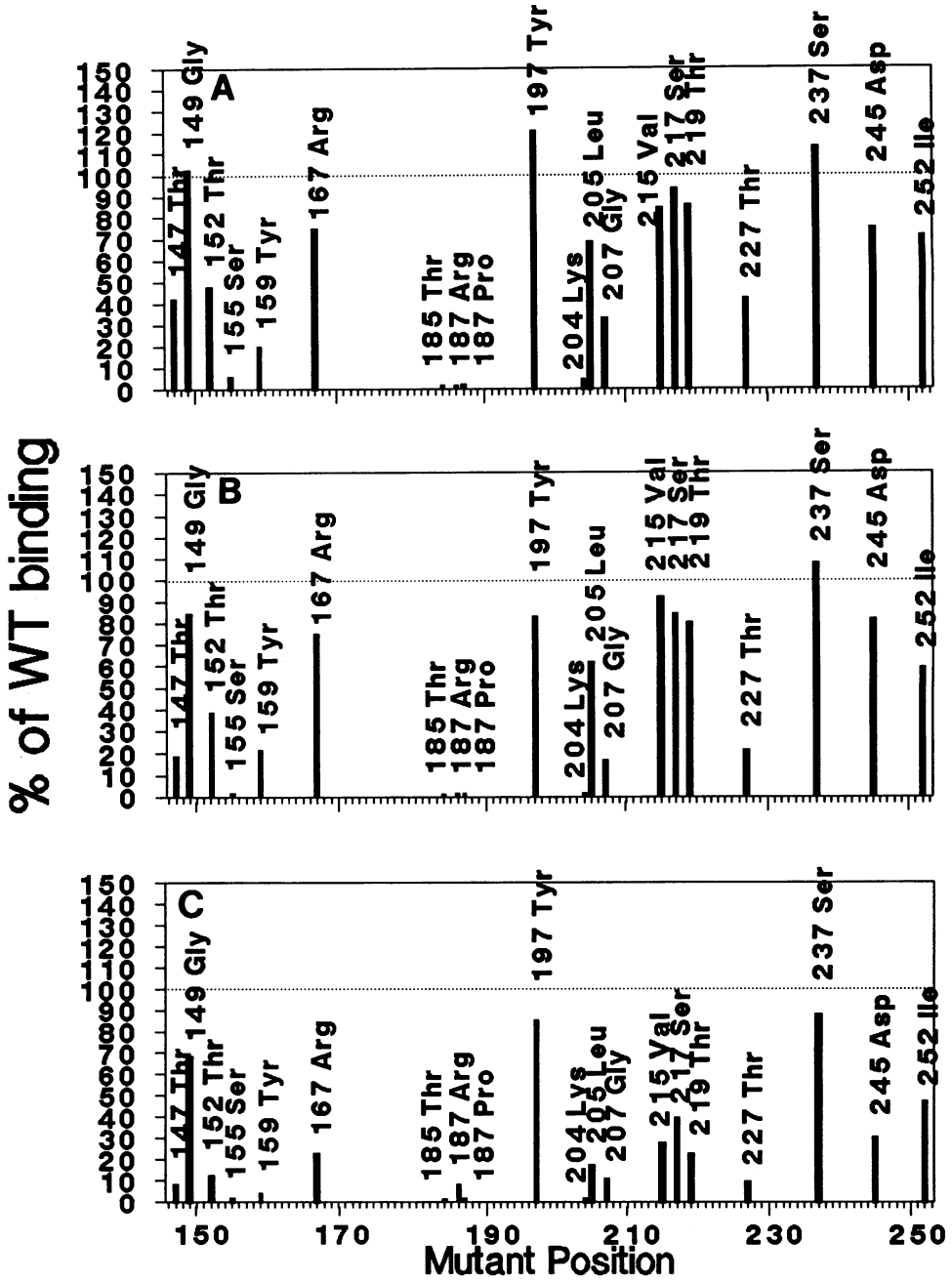

FIG. 4. Binding of mutant T antigens to sites I and II at the replication origin. Immunoprecipitated mutant T antigens were tested in a binding assay with a labeled fragment containing the wild-type SV40 origin (A), site I DNA (B) (*Taq*I fragment E of pOS1), or site II DNA (C) (*Taq*I fragment E of pSVOdl3). All reactions were performed with a 1,000-fold excess of competitor DNA. Results are presented as a percentage of wild-type T-antigen binding under each condition.

The results of our experiments implicate multiple regions within the DNA-binding domain of SV40 T antigen in the efficient binding to various substrates. By the criteria described above, two regions appear to be of primary importance in binding to DNA at sites I and II in a sequence-specific fashion. These regions are illustrated by the DNA-binding properties of the proteins with mutations mapping at residues 155 and 204 (and to a smaller extent at residue 207). Of all the proteins tested, the first two mutants showed the greatest percent inhibition of binding to *ori* DNA when tested in the presence of competitor calf thymus DNA (Fig. 3 and Table 1). The boundaries of these two regions can be deduced from the effects of neighboring mutations on *ori* binding (Fig. 3 and 4), from the pattern of replicating and nonreplicating mutants (Fig. 1), from DNA-binding data for T-antigen mutants described in the literature (2, 24, 34) and from the comparison of amino acid sequences of various papovavirus T antigens (compiled by J. Pipas). The first such region, which we call A1 (Fig. 6), most likely extends from residues 152 to 155 (Ser-Asn-Arg-Thr). This 4-amino-acid sequence is strongly conserved among all papovavirus T antigens (Fig. 6). This is significant because at least some of

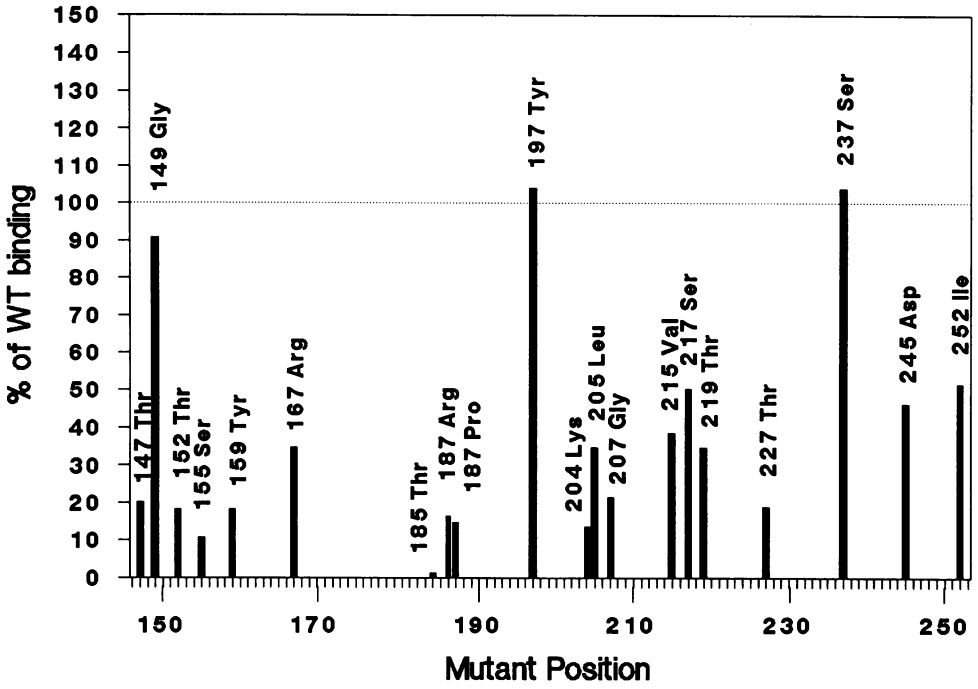

FIG. 5. Binding of mutant T antigens to site II DNA under replication conditions. Immunoprecipitated mutant T antigens were tested in a binding assay with labeled site II DNA under replication conditions. All reactions included ATP and competitor calf thymus DNA and were performed at 37°C.

those proteins (those from SV40, BK virus, JC virus, and polyomavirus) can bind to the same viral origin (15, 33, 35), and all probably recognize the same pentanucleotide sequence GAGGC (12, 13, 33). Two of the four residues (152 and 155) in this region have been mutated in our study, and these mutants are deficient in binding to sites I and II (Fig. 4 and Table 1) but not to plasmid DNA (Fig. 2). A third residue (153) has been implicated in DNA binding by Prives et al. (34). Furthermore, this sequence lies within a larger region (from about residue 144 to 158) identified by Paucha et al. (32) which appeared to be important in sequence-specific binding. We can take a good guess at the limits of this region because sequence homology among various papovavirus T antigens is considerably weaker immediately downstream and to some extent upstream of A1. Second, a mutation at residue 149 had no effect on *ori* binding (Fig. 3 and 4), and a mutation at residue 157 had no effect on virus replication (Fig. 1). It is worth noting that region A1 contains a potential N-glycosylation site.

By similar reasoning, a second region (called B2) appears to be critical for recognition of the origin. This second sequence extends from residue 203 to 207 (Fig. 6) and is just as important for *ori* binding as region A1. The 5-amino-acid sequence of region B2 is also highly conserved among all six papovaviruses (Fig. 6). Sequence similarity drops significantly outside of this region, and mutations at residues 202 and 209 had no effect on virus replication (Fig. 1). Three of the five residues (204, 205, and 207) in B2 were mutated in this study and appear to be important for binding to site II (205) or to both sites I and II (204 and 207) (Fig. 3 and 4, Table 1). (Subsequent analysis of the residue 205 mutant showed it to be partially defective in binding to site I as well.) Furthermore, a residue 203 (His to Gln) mutant is defective in origin-specific binding (24).

It is important to note that not all amino acids within regions A1 and B2 appear to have equal importance in binding to the origin. For example, in region B2 of the wild-type protein, Arg-204 may be much more important than Ala-207 or Val-205, although mutations at all three sites show an effect. Certain residues within elements A1 and B2, and in particular Thr-155 and Arg-204, may therefore contribute more to the direct contact with DNA sequences at the origin. Additional sites are most likely involved in making the proper contacts with the origin. Saturation mutagenesis of these regions coupled with the substitution of various amino acids at each site should provide useful information in this regard.

A third region, termed B1, is also crucial for the interaction with DNA (Fig. 6). This region, unlike A1 and B2, appears to be important in the general recognition of DNA and not strictly in sequence-specific binding. There are two major reasons for believing that this is the case. First, a mutation at residue 185 or 187 seriously reduces DNA-binding activity to all substrates tested, including plasmid DNA. Second, region B1 (residues 183 to 187) is identical in the T antigens of SV40, BK virus, and JC virus, but the corresponding sequence is strikingly different in the other papovaviruses except at residue 183 (Fig. 6). This implies that B1 is present in this region only in the first three virus proteins. In the other three, (B-lymphotropic virus, hamster virus, and polyomavirus) B1 is either absent or its functional equivalent is present elsewhere in the DNA-binding domain. The limits of this region were placed at 183 and 187 because mutants with mutations at residues 182 and 189 replicate (Fig. 1) and residue 188 is not conserved in SV40, BK virus, and JC virus. Since the mutants with changes in this region were unable to bind to site I or II in addition to being defective in nonspecific DNA binding (Fig. 3 and 4), it suggests that T antigen must bind DNA nonspecifically first before it can efficiently interact with origin sequences.

We hypothesize, therefore, that the critical sequences for origin binding make up a motif of three major elements and

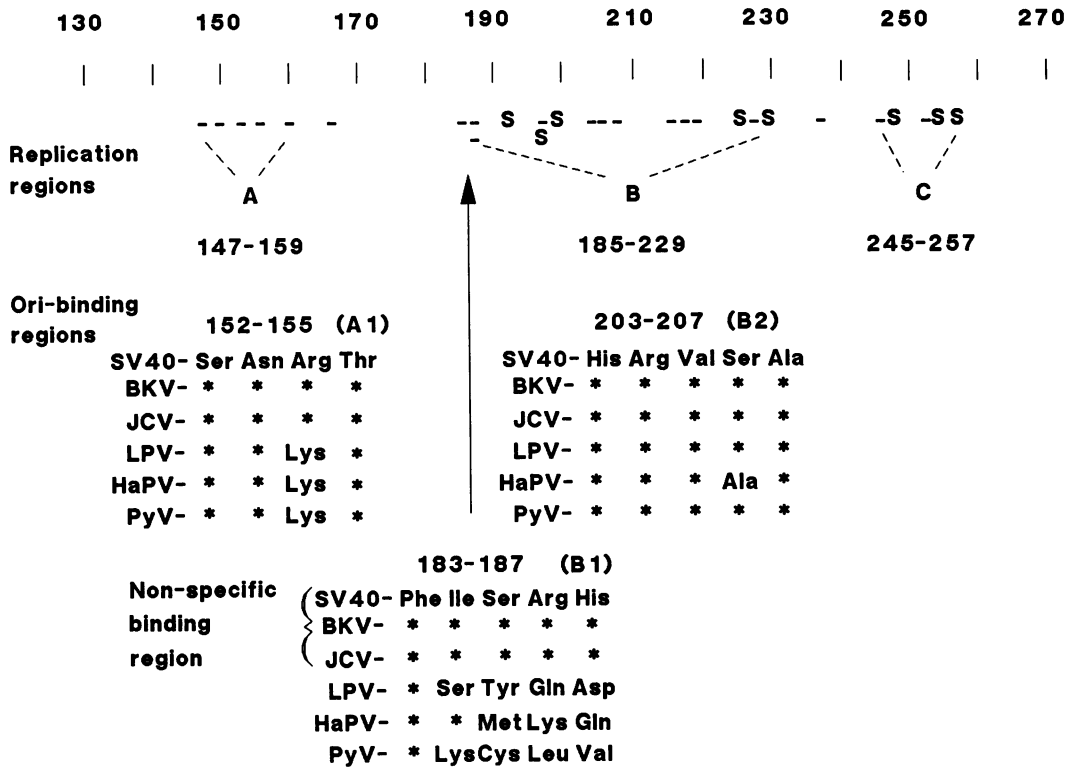

FIG. 6. Sequence elements of the *ori*-binding motif. The *ori*-binding motif is thought to be composed of three sequence elements. Two of the three elements (A1 and B2) are necessary for the proper recognition and binding to sites I and II at the origin. The third element (B1) is important for nonspecific binding to DNA. Shown are the amino acid sequences that correspond to these three elements for SV40, BK virus (BKV), JC virus (JCV), lymphotropic papovavirus (LPV), hamster papovavirus (HaPV), and polyomavirus (PyV) T antigens. Note that the central element B1 is present in the first three virus T antigens only. The equivalent element in LPV, HaPV, and PyV may be present elsewhere in the protein.

that the central element (B1) is necessary for nonspecific DNA interactions. An appealing possibility is that some residues in B1 (as well as other sequences) are involved in making the initial contact with the DNA. Anchored in part through B1, the DNA slides along until the GAGGC pentanucleotide sequences appear, at which point elements A1 and B2 (or certain residues within them) also make contact.

In addition to the three motif elements described above, at least one region (region B3 [residues 215 to 219]) is important for the proper recognition of site II sequences. Three mutants with mutations within this region show altered binding to site II DNA only (Fig. 3). Other potential regions involve sequences surrounding residues 167 (166 and 167?) and 245 (245 to 247?) (Fig. 7). At the moment, it is difficult to explain how certain protein sequences could be important in binding to site II but not site I DNA, since both sites contain very similar pentanucleotide sequences that are recognized by T antigen (12). Perhaps the organization of the pentanucleotides or the general structure of the DNA itself is important. Site I contains two GAGGC pentanucleotide segments oriented towards the early region, and site II contains four such segments, two oriented one way and two oriented the other way (12). The structure of the DNA at these two sites could also be a determining factor, since the DNA appears to be bent at site I (36) but not at the four pentanucleotides of site II, although bending occurs in the AT tract adjacent to them (9). Another difference is the fact that T antigen probably makes additional contacts within the inverted repeat sequences in site II (4), a DNA element that is not present in site I. Whatever the reasons, it appears that further protein-DNA contacts are important for the proper binding to site II, which in turn is required for the initiation of virus DNA replication (29, 38). Our data suggest that element B3 in T antigen (Fig. 7) and perhaps other sites as well (residues 167 and 245) are involved in these interactions.

The four regions described above appear to be the most important for proper recognition and binding to various origin sequences, but they are certainly not the only regions involved. It is likely that residues 147, 159, 225 to 229, and 252 make some contribution to the proper structure and function of the *ori*-binding domain (Fig. 7). Unlike the amino acids in each of the four major elements described above, it is not possible to predict whether these residues are important for proper conformation or whether they have a direct functional role. This is either because the mutations are isolated or because the magnitude of the effect on DNA binding was small. On the basis of the phenotypic effects of its mutation, residue 159 could, like element B1, be important for nonspecific DNA binding.

Our results may explain the phenotype of a mutant described by Margolskee and Nathans (26), which had a change at residue 157 (Ala to Leu). This mutant was isolated as a second-site revertant of another mutation in one of the four GAGGC pentanucleotides in site II. The mutant T antigen displayed a relaxed specificity for the origin in that it recognized the wild-type as well as the altered origins. Given its proximity to element A1, it is conceivable that the mutation at residue 157 altered the local conformation and indirectly affected the function of this element or perhaps of

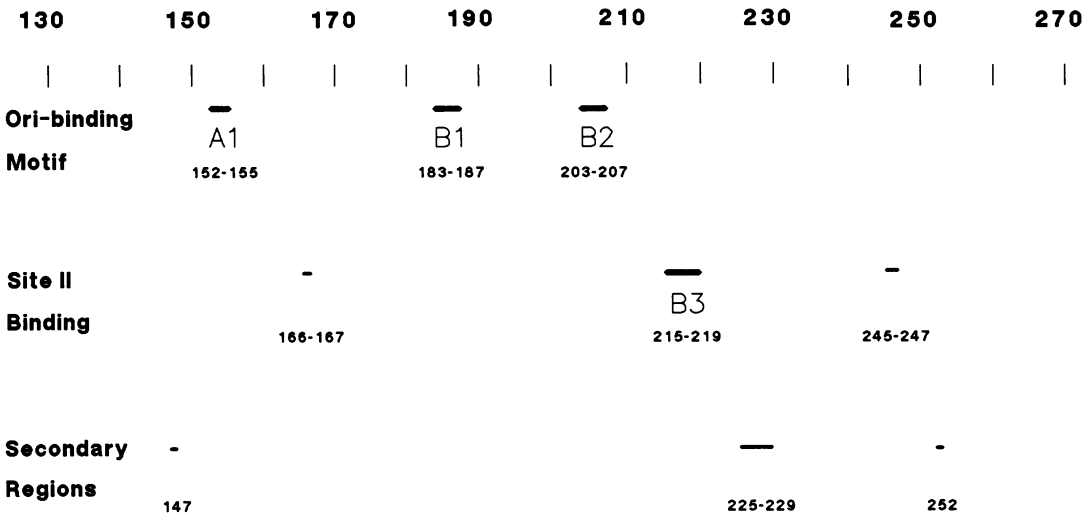

FIG. 7. Map of important regions in the DNA-binding domain. Three categories of elements are shown. At the top is the tripartite motif important to origin binding, in the middle are three regions that appear to be required for the proper binding to site II, and at the bottom are several secondary regions which have a lesser influence on the activity of the domain.

element B2 (if they are close together in space). However, only three-dimensional data can directly address this issue.

Our analysis most likely underestimates the number of important sites, since not all positions have been mutated, and the possibility exists that, at some sites where mutations did not result in a phenotypic change, a different amino acid substitution may show an effect. Furthermore, since our DNA-binding assays were performed at protein excess and were not under equilibrium conditions, a greater effect on DNA binding may have been observed in some cases. On the whole, however, it seems likely that the overall pattern of contributing sites is not very different from that shown in Fig. 1 as regions A, B, and C, although this figure gives no indications of the relative importance of each residue to domain function.

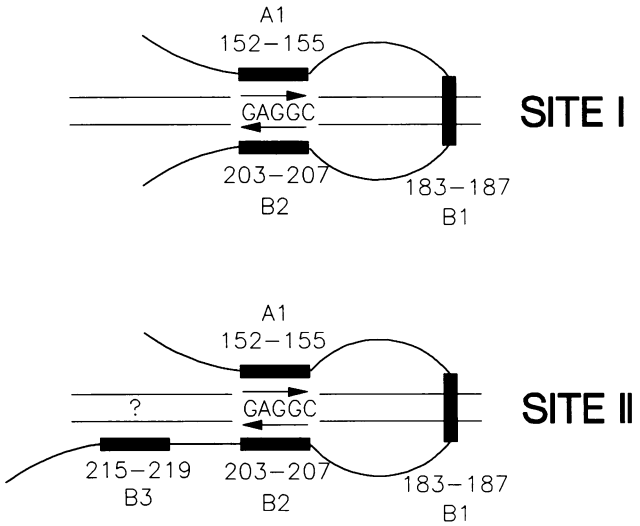

FIG. 8. Basic model of DNA interactions. This model describes the involvement of various sequence elements in binding to sites I and II at the origin. At site I, the major elements are A1, B1, and B2. A1 and B2 interact with the GAGGC pentanucleotides, whereas B1 recognizes a general structural feature of DNA. The arrows on either side of the GAGGC sequence signify that the orientation of the DNA relative to the DNA-binding domain is unknown. At site II, the same three elements are involved, but at least one (B3) and possibly two others are also required for binding. The sites of contact of these latter elements on the DNA are not known.

The proper structures of the regions directly involved in DNA binding are of obvious importance. There is evidence that the structure of element B1 is crucial, since a mutant which maps at residue 186 (Arg to Thr) is temperature-sensitive (19), implying that a structural change in the protein which makes it incapable of recognizing DNA at the nonpermissive temperature has occurred. Secondary structure programs based on the methods of Garnier et al. (16) and Novotny and Auffray (30) predict that the sequence in B1 (183 to 187) or part of it forms a β-turn, analogous to the β-turn in the helix-turn-helix motifs of certain DNA-binding proteins. However, no clear predictions can be made for the secondary structures A1 and B2, the sequence-specific elements. Predictions from various programs are compatible with a β-sheet, a β-turn, or an extended conformation. Perhaps the only consistency is the fact that none predict an α-helix for either A1 or B2.

Our major points are illustrated in a simple working model that shows the involvement of various sequence elements in binding to sites I and II (Fig. 8). At site I, a tripartite motif consisting of elements A1, B1, and B2 is associated with the DNA. A1 and B2 make sequence-specific contacts with the GAGGC pentanucleotides. Since these two elements recognize the same DNA sequences, they could be close together in three-dimensional space. Region B1 may consist of a β-turn and binds to DNA nonspecifically. At site II, the same motif is involved, but in addition, element B3 and possibly others make contact with some unknown DNA sequences. Not shown, but implied, is the fact that these sites of contact are only possible in the proper structural background of the DNA-binding domain.

ACKNOWLEDGMENTS

We thank Mary Anderson, William Young, and Hema Mallick for expert technical assistance and Peter Kissel and Jacques Pene for preparing the oligonucleotides. We also acknowledge Pat Hearing and his students for helping us set up the adenovirus expression

system and Alan Smith for pointing out the N-glycosylation site in region A1.

This work was supported by an RSA (CA 08466) from the Public Health Service to D.T.S. and by Public Health Service grants CA 36118, CA 18808, and CA28146 from the National Cancer Institute. This work was initiated while D.T.S. was on sabbatical leave at Stony Brook.

LITERATURE CITED

1. **Arthur, A. K., A. Höss, and E. Fanning.** 1988. Expression of simian virus 40 T antigen in *Escherichia coli*: localization of T-antigen origin DNA-binding domain to within 129 amino acids. J. Virol. **62**:1999–2006.
2. **Auborn, K., M. Guo, and C. Prives.** 1989. Helicase, DNA-binding, and immunological properties of replication-defective simian virus 40 mutant T antigens. J. Virol. **63**:912–918.
3. **Berkner, K. L., and P. A. Sharp.** 1983. Generation of adenovirus by transfection of plasmids. Nucleic Acids Res. **11**:6003–6020.
4. **Borowiec, J., and J. Hurwitz.** 1988. Localized melting and structural changes in the SV40 origin of replication induced by T-antigen. EMBO J. **7**:3149–3158.
5. **Borowiec, J., and J. Hurwitz.** 1988. ATP stimulates the binding of simian virus 40 (SV40) large tumor antigen to the SV40 origin of replication. Proc. Natl. Acad. Sci. USA **85**:64–68.
6. **Chen, C., and H. Okayama.** 1987. High-efficiency transformation of mammalian cells by plasmid DNA. Mol. Cell. Biol. **7**:2745–2752.
7. **Clark, R., K. Peden, J. M. Pipas, D. Nathans, and R. Tjian.** 1983. Biochemical activities of T-antigen encoded by simian virus 40 A gene deletion mutants. Mol. Cell. Biol. **3**:220–228.
8. **Dean, F. B., P. Bullock, Y. Murakami, R. Wobble, L. Weissbach, and J. Hurwitz.** 1987. Simian virus 40 (SV40) DNA replication: SV40 large T antigen unwinds DNA containing the SV40 origin of replication. Proc. Natl. Acad. Sci. USA **84**: 16–20.
9. **Deb, S., A. L. DeLucia, A. Koff, S. Tsui, and P. Tegtmeyer.** 1986. The adenine-thymine domain of the simian virus 40 core origin directs DNA bending and coordinately regulates DNA replication. Mol. Cell. Biol. **6**:4578–4584.
10. **Deb, S., S. Tsui, A. Koff, A. L. DeLucia, R. Parsons, and P. Tegtmeyer.** 1987. The T-antigen-binding domain of the simian virus 40 core origin of replication. J. Virol. **61**:2143–2149.
11. **Deb, S. P., and P. Tegtmeyer.** 1987. ATP enhances the binding of simian virus 40 large T antigen to the origin of replication. J. Virol. **61**:3649–3654.
12. **DeLucia, A. L., B. A. Lewton, R. Tjian, and P. Tegtmeyer.** 1983. Topography of simian virus 40 A protein-DNA complexes: arrangement of pentanucleotide interaction sites at the origin of replication. J. Virol. **46**:143–150.
13. **Dilworth, S., A. Cowie, R. Kamen, and B. Griffin.** 1984. DNA-binding activity of polyoma virus large tumor antigen. Proc. Natl. Acad. Sci. USA **81**:1941–1945.
14. **Dodson, M., F. B. Dean, P. Bullock, H. Echols, and J. Hurwitz.** 1987. Unwinding of duplex DNA from the SV40 origin of replication by T antigen. Science **238**:964–967.
15. **Frisque, R. J.** 1983. Regulatory sequences and virus-cell interactions of JC virus, p. 41–59. *In* J. L. Sever and D. L. Madden (ed.), Polyomaviruses and human neurological disease. Alan R. Liss, Inc., New York.
16. **Garnier, J., D. J. Osguthorpe, and B. Robson.** 1978. Analysis of the accuracy and implication of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. **120**:97–120.
17. **Gottlieb, P., M. S. Nasoff, E. F. Fisher, A. M. Walsh, and M. H. Caruthers.** 1985. Binding studies of SV40 T-antigen to SV40 binding site II. Nucleic Acids Res. **13**:6621–6634.
18. **Harlow, E., L. V. Crawford, D. C. Pim, and N. M. Williamson.** 1981. Monoclonal antibodies specific for simian virus 40 tumor antigen. J. Virol. **39**:861–869.
19. **Hutchinson, N. I., L.-S. Chang, M. M. Pater, N. Bouck, T. E. Shenk, and G. DiMayorca.** 1985. Characterization of a new simian virus 40 mutant, *tsA*3900, isolated from deletion mutant *tsA*1499. J. Virol. **53**:814–821.
20. **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.
21. **Jones, N., and T. Shenk.** 1979. Isolation of adenovirus type 5 host range deletion mutants defective for transformation of rat embryo cells. Cell **17**:683–689.
22. **Kunkel, T. A.** 1985. Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc. Natl. Acad. Sci. USA **82**:488–492.
23. **Loeber, G., R. Parsons, and P. Tegtmeyer.** 1989. The zinc finger region of simian virus 40 large T antigen. J. Virol. **63**:94–100.
24. **Manos, M. 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.
25. **Mansour, S. L., T. Grodzicker, and R. Tjian.** 1985. An adenovirus vector system used to express polyoma virus tumor antigens. Proc. Natl. Acad. Sci. USA **82**:1359–1363.
26. **Margolskee, R. F., and D. Nathans.** 1984. Simian virus 40 mutant T antigens with relaxed specificity for the nucleotide sequence at the viral DNA origin of replication. J. Virol. **49**:386–393.
27. **McCutchan, J. H., and J. S. Pagano.** 1968. Enhancement of the infectivity of simian vius 40 deoxyribonucleic acid with diethylaminoethyldextran. J. Natl. Cancer Inst. **41**:351–357.
28. **McKay, R.** 1981. Binding of simian virus 40 T antigen-related protein to DNA. J. Mol. Biol. **145**:471–488.
29. **Myers, R. M., 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.
30. **Novotny, J., and C. Auffray.** 1984. A program for prediction of protein secondary structure from nucleotide sequence data: application to histocompatibility antigens. Nucleic Acids Res. **12**:243–255.
31. **Pakula, A. A., and R. T. Sauer.** 1989. Genetic analysis of protein stability and function. Annu. Rev. Genet. **23**:289–310.
32. **Paucha, E., D. Kalderon, R. W. Harvey, and A. E. Smith.** 1986. Simian virus 40 origin DNA-binding domain on large T antigen. J. Virol. **57**:50–64.
33. **Pomeranz, B., and J. Hassel.** 1984. Polyomavirus and simian virus 40 large T antigens bind to common DNA sequences. J. Virol. **49**:925–937.
34. **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.
35. **Ryder, K., A. L. DeLucia, and P. Tegtmeyer.** 1983. Binding of SV40 A protein to the BK virus origin of DNA replication. Virology **129**:239–245.
36. **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.
37. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
38. **Shortle, D., and D. Nathans.** 1979. Regulatory mutants of simian virus 40: constructed mutants with base substitutions at the origin of DNA replication. J. Mol. Biol. **131**:801–817.
39. **Simmons, D. T.** 1986. DNA-binding region of the simian virus 40 tumor antigen. J. Virol. **57**:776–785.
40. **Simmons, D. T.** 1988. Geometry of the simian virus 40 large tumor antigen-DNA complex as probed by protease digestion. Proc. Natl. Acad. Sci. USA **85**:2086–2090.
41. **Simmons, D. T., W. Chou, and K. Rodgers.** 1986. Phosphorylation down regulates the DNA-binding activity of simian virus 40 T antigen. J. Virol. **60**:888–894.
42. **Solnick, D.** 1983. Shuffling adenovirus promoters: a viral recombinant with early region 1A under late transcriptional control. EMBO J. **2**:845–851.
43. **Stahl, H., P. Droege, and R. Knippers.** 1986. DNA helicase activity of SV40 large tumor antigen. EMBO J. **5**:1939–1944.
44. **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.
45. **Stow, N. D.** 1981. Cloning of a DNA fragment from the left-hand

terminus of the adenovirus type 2 genome and its use in site-directed mutagenesis. J. Virol. **37**:171–180.

46. **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.
48. **Tjian, R.** 1978. Protein-DNA interactions at the origin of simian virus 40 DNA replication. Cold Spring Harbor Symp. Quant. Biol. **43**:655–662.
48. **Tjian, R.** 1978. The binding site of SV40 DNA for a T antigen-related protein. Cell **13**:165–179.
49. **Wold, M. S., J. J. Li, and T. J. Kelly.** 1987. Initiation of simian virus 40 DNA replication in vitro: large-tumor-antigen and origin-dependent unwinding of the template. Proc. Natl. Acad. Sci. USA **84**:3643–3647.