Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 2086–2090, April 1988
Biochemistry

Geometry of the simian virus 40 large tumor antigen–DNA complex as probed by protease digestion

(DNA–protein complexes/DNA-binding domain/protease cleavage/DNA-cellulose chromatography)

DANIEL T. SIMMONS

School of Life and Health Sciences, University of Delaware, Newark, DE 19716

Communicated by Norman Davidson, November 30, 1987 (received for review August 20, 1987)

ABSTRACT **I have mapped the regions of simian virus 40 (SV40)-encoded large tumor (T) antigen that remained associated with origin-containing SV40 DNA after digestion with various concentrations of Pronase E, a nonspecific protease. Immunoaffinity-purified, labeled T antigen was bound to SV40 DNA-cellulose and treated with Pronase E. A "core" region representing amino acids 140 to about 281 was identified by peptide mapping in the fraction that remained bound to the DNA-cellulose after treatment with a high (135 μg/ml) concentration of enzyme. This region corresponds to the DNA-binding domain of the protein molecule. After treatment with Pronase E at 66 μg/ml, the bound fraction consisted of the DNA-binding domain and a region that extends to residue 371. This larger protein segment binds more stably to the viral DNA than does the core by itself. At lower concentrations of Pronase E, additional sequences from the NH_2-terminal region of T antigen and from the COOH-half of the molecule were observed in the bound fractions. Linear maps of resistant regions, generated for each concentration of protease used, provide information about the geometry of the protein molecule associated with the DNA. I suggest that regions that are easily cleaved by the protease are exposed in the DNA–protein complex, whereas those that remain bound to the DNA at increasing concentrations of the enzyme represent segments that are in progressively closer proximity to the viral DNA origin.**

The mechanism by which simian virus 40 (SV40) large tumor (T) antigen binds to DNA is not understood. Two protein structures have been described in the literature that correspond to DNA-binding domains. The first, a DNA-binding motif consisting of helix–turn–helix, is very common in a number of sequence-specific DNA-binding proteins including the bacteriophage λ Cro and repressor proteins and the bacterial catabolite gene-activator protein (CAP) (1). Specific α-helices of these proteins interact with the major groove of the DNA. Second, DNA binding has been associated with a structure called a "Zn^{2+} finger," which consists of two cysteine and two histidine residues separated by a loop of up to 30 residues and binding a centrally located zinc ion (2–4). Two or more Zn^{2+} fingers are thought to be involved in the binding to the DNA.

The DNA-binding region of SV40 T antigen has been identified by genetic (5–7) and biochemical (8) methods. The central region probably starts near residue 140 (6) and extends to about 259 (9), although other important sequences may extend to 371 (5, 8). No apparent sequence similarity with prokaryotic helix–turn–helix DNA-binding proteins has been detected in this region, but a putative Zn^{2+} finger exists between residues 302 and 320 (2).

The DNA-binding domain recognizes the consensus pentanucleotide sequence GAGGC, which is present in multiple copies at and around the origin of replication in SV40 DNA (10–13). Binding to all four of these pentanucleotides in T-antigen binding region II is required for T antigen to initiate the replication of viral DNA (14). Hence, four molecules of T antigen are involved in DNA replication. These protein molecules interact with the major groove of the DNA (11), and it is reasonable to speculate that protein–DNA and protein–protein interactions are necessary for the proper recognition of the origin.

In order to better understand the interaction between T antigen and SV40 DNA, the regions of the protein molecule that remain bound to the DNA after protease digestion have been mapped. The aim of the work described here was to identify the regions of T antigen that make strong, intermediate, and weak associations with the viral origin, and consequently to obtain some information about the spatial relation between T antigen and the viral DNA. The method described may be useful in the study of other DNA-binding proteins.

MATERIALS AND METHODS

Gel Retardation Assays. SV40 T antigen was purified by immunoaffinity chromatography using PAb419 monoclonal antibody essentially as described by Simanis and Lane (15). The T antigen was extracted from Ad-SVR284-infected HeLa cells (16) as previously described (17). pSVO+ plasmid DNA, which contains a wild-type SV40 origin region (18), was labeled with ^{32}P by nick-translation and cleaved with restriction endonuclease *Taq* I. Fragment E (about 300 base pairs), which contains the origin, was purified by agarose gel electrophoresis and incubated with sufficient T antigen to completely bind the DNA fragment. The binding reaction was allowed to proceed in DNA binding buffer [20 mM NaH_2PO_4/0.15 M NaCl/0.1 mM EDTA/0.05% Nonidet P-40/3% (vol/vol) dimethyl sulfoxide/10% (vol/vol) glycerol] for 1 hr at 23°C in the presence of competitor calf thymus DNA (5 μg/ml). In some samples, the complex was digested with Pronase E (Sigma protease type XIV, preincubated for 1 hr at 37°C) for an additional hour at 23°C. This was immediately followed by electrophoresis in a 5% acrylamide gel as described (19).

DNA-Cellulose Chromatography. DNA-cellulose was prepared exactly as described (8). A mutant SV40 DNA (ev1103) containing five to seven replication origins in tandem was used to prepare the multiorigin SV40 DNA-cellulose. All DNA columns contained a ratio of 1.5 mg of DNA per gram of cellulose powder.

Immunoaffinity-purified T antigen labeled *in vivo* with [^{35}S]methionine was passed several times through a 1-cm column in DNA binding buffer at 4°C. The column was washed with about 50 ml of DNA binding buffer and the cellulose was incubated with 3 ml of Pronase E in DNA

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: SV40, simian virus 40; T antigen, SV40-encoded large tumor antigen.

binding buffer at 23°C for 1 hr with occasional mixing of the resin. The cellulose was then washed with 20 ml of DNA binding buffer, and T-antigen fragments that were still attached were eluted with 0.01 M Tris, pH 8.0/1 M NaCl/0.05% Nonidet P-40. Fragments were precipitated with 20% trichloroacetic acid in the presence of bovine albumin (20 μg/ml) as carrier.

Tryptic Peptide Mapping and Gel Electrophoresis. Trichloroacetic acid-precipitated fragments were washed twice with cold acetone and twice with cold absolute ethanol. For peptide analysis, the fragments were treated with performic acid, lyophilized, and digested with trypsin as described (20). Tryptic peptides were separated in two dimensions on thin-layer cellulose plates as described (21). For gel electrophoresis, the fragments were resuspended in gel sample buffer and subjected to $NaDodSO_4$/PAGE as described (8).

RESULTS

Digestion of T-Antigen–DNA Complexes with Pronase E. I first determined the proper conditions for cleaving the exposed parts of the T-antigen molecule in the DNA–protein complex while preserving the association of the binding region with the DNA. Immunoaffinity-purified unlabeled T antigen was incubated with a ^{32}P-labeled SV40 DNA origin fragment and then digested with various concentrations of Pronase E, a nonspecific protease. The sizes of the DNA–protein complexes were analyzed by electrophoresis in nondenaturing 5% acrylamide gels (Fig. 1). The uncleaved T-antigen–DNA complex migrated very close to the top of the gel (A in Fig. 1, lane 2), but after Pronase E treatment, the complex moved substantially into the gel (B in Fig. 1, lanes 3 and 4), although not as far as free DNA (Fig. 1, lane 1). At high concentrations of Pronase E (Fig. 1, lanes 5 and 6), the complex fell apart. Although similar observations were made when other proteases (notably *Staphylococcus aureus* V8 protease and chymotrypsin) were used, Pronase E gave the most consistent results. It was thought that the intermediate concentrations of Pronase E used in these experiments might be useful in mapping the regions of T antigen that are still attached to the DNA after protease digestion.

Specific Regions of T Antigen Remain Bound to SV40 DNA after Pronase E Digestion. The most effective approach to map DNA-connected regions of T antigen was by DNA-cellulose chromatography. T antigen was labeled *in vivo* with [^{35}S]methionine and purified by immunoaffinity. Multiorigin-containing SV40 DNA was bound to purified cellulose, where it served as an affinity matrix for T antigen. The labeled T antigen was passed over the columns; the columns were extensively washed and the cellulose was incubated with Pronase E at concentrations determined in the gel retardation experiment. After a subsequent wash step, any T-antigen fragments that were still bound to the columns were eluted with buffer containing 1 M NaCl. The origin of the fragments on T antigen was determined by two-dimensional fingerprinting of tryptic peptides on thin-layer cellulose plates (21). This identification was possible because previous work had provided a detailed linear map of methionine-containing tryptic peptides of this protein (8).

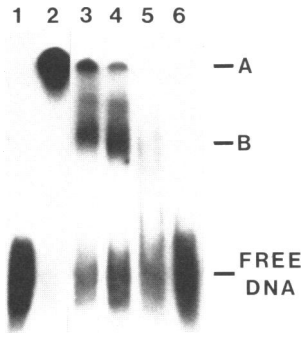

FIG. 1. Gel retardation assays of Pronase E-treated SV40 T-antigen–DNA complexes. About 2 μg of unlabeled immunoaffinity-purified SV40 T antigen was mixed with a ^{32}P-labeled DNA fragment containing the origin for DNA replication. Samples were treated with various concentrations of Pronase E and applied to a nondenaturing 5% acrylamide gel. Lane 1, no T antigen (control); lane 2, complexes incubated without Pronase E; lanes 3–6, complexes incubated with Pronase E at 20, 66, 200, or 1000 μg/ml, respectively. "A" and "B" mark the positions of the complex of DNA with intact T antigen and proteolyzed T antigen, respectively.

It was first important to show that the Pronase E-resistant fragments that were eluted from the column with high salt were due to the association of the protein with the origin region of the viral DNA. Consequently, this experiment was performed with a control column not containing any DNA, with a column containing calf thymus DNA, and with a multiorigin SV40 DNA column. Fig. 2 (1 M NaCl lanes) shows that no fragments were eluted from the DNA-free column after Pronase E digestion and that although a few fragments could be recovered from the column containing calf thymus DNA, substantially more labeled fragments were bound to and could be eluted from the SV40 DNA column. It was concluded that the majority of these eluted fragments originated from T antigen that had bound specifically to the viral replication origin.

It should be noted that the T-antigen fragments that were retained on the SV40 DNA-cellulose column after the Pronase E step (Fig. 2, 1 M NaCl lanes) did not appear in the fraction that was washed from the column (Fig. 2, wash lanes). Consequently, the 38-, 30-, and 14-kDa fragments, which were found mainly in the high-salt eluate, represent a unique class of fragments that were specifically retained on the DNA-cellulose.

Mapping of DNA-Associated Regions. In order to map the DNA-associated regions of the protein, [^{35}S]methionine-labeled T antigen was bound to multiorigin SV40 DNA cellulose columns and the bound protein was treated with

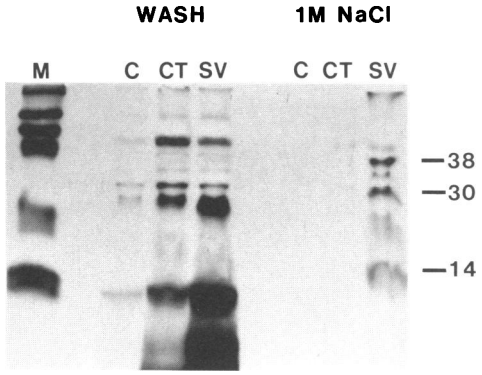

FIG. 2. Identification of T-antigen fragments specifically bound to SV40 DNA-cellulose after protease digestion. [^{35}S]Labeled T antigen was passed through a control (no DNA) cellulose column (C), a cellulose column containing calf thymus DNA (CT), or a column containing multiorigin SV40 DNA (SV). The columns were washed extensively and the cellulose was incubated with Pronase E (66 μg/ml) for 1 hr at 23°C. Fragments that were released from the columns under those conditions were recovered by trichloroacetic acid precipitation and applied to the gel (wash lanes). After additional wash steps, the fragments that remained bound to the columns were eluted with 1 M NaCl-containing buffer, precipitated with trichloroacetic acid, and applied to the gel (1 M NaCl lanes). Numbers at right indicate molecular mass (in kDa) of eluted fragments, determined by comparison with marker proteins (lane M, from top to bottom) of 200, 97, 68, 43, 26, 18, and 14 kDa.

various amounts of Pronase E. After a subsequent wash to remove unbound fragments, the bound fragments were eluted with high-salt buffer and treated with trypsin, and the resulting tryptic peptides were separated in two dimensions on thin-layer plates. No effort was made to first separate the fragments from one another on acrylamide gels, since it was important to analyze the entire fraction. However, compatible results were obtained when specific gel-purified fragments were analyzed (data not shown).

The fingerprints of the eluted fragments are shown in Fig. 3. When the T-antigen–DNA complexes were treated with Pronase E at 135 μg/ml, the bound region yielded a simple fingerprint consisting of T-antigen peptides 15, 19, and 20 (see Fig. 4). A nearly identical fingerprint was obtained from a 16-kDa tryptic fragment (8) that maps from residue 131 to 281. With Pronase E at 66 μg/ml, the high-salt fraction generated a fingerprint that is essentially identical to that of a tryptic 30-kDa fragment mapping from 131 to 371 (8). As the concentration of Pronase E was lowered, additional sequences could be identified in the bound fractions. In particular, with Pronase E at 33 μg/ml, three peptides from the NH_2-terminal region (M, N, and 3; see Fig. 4) were noted. Peptides M and N are the first two methionine-containing tryptic peptides of T antigen and lie within the first 21 amino acids. Peptide 3 is Met-Lys (22) and is found in two different places in the protein chain (residues 52–53 in the NH_2-terminal region and residues 615–616 in the COOH-terminal region). Since the COOH-terminal end of T antigen is missing from all of these protected fractions (see below), it is very unlikely that peptide 3 in this sample originated from residues 615–616. After treatment with the enzyme at 10 μg/ml, the fingerprint contained two additional peptides (S and 18) from the NH_2-terminal region, corresponding to residues 68–130 and possibly to residues 62–67, respectively. At still lower concentrations, the DNA-associated fraction also contained sequences from the middle of the molecule (peptide 8, see Fig. 4) and further downstream (peptide 5).

The summary of these results is depicted in Fig. 4. A map of the methionine-containing tryptic peptides of T antigen is shown at the top of the figure (8). Also shown are maps of three previously studied tryptic fragments to which the high-salt-eluted fragments are compared. The regions that are still bound to the DNA at various concentrations of pronase E are shown in the lower part of the figure.

DISCUSSION

This work has identified a "core" region of SV40 T antigen that remains bound to SV40 DNA in the presence of Pronase E at 135 μg/ml. This core region most likely represents a 14-kDa fragment that is observed in the resistant fraction at this concentration of enzyme (data not shown, but see Fig. 2). The fragment is a little smaller than a 16-kDa tryptic fragment that maps at residues 131–281. Based on the size

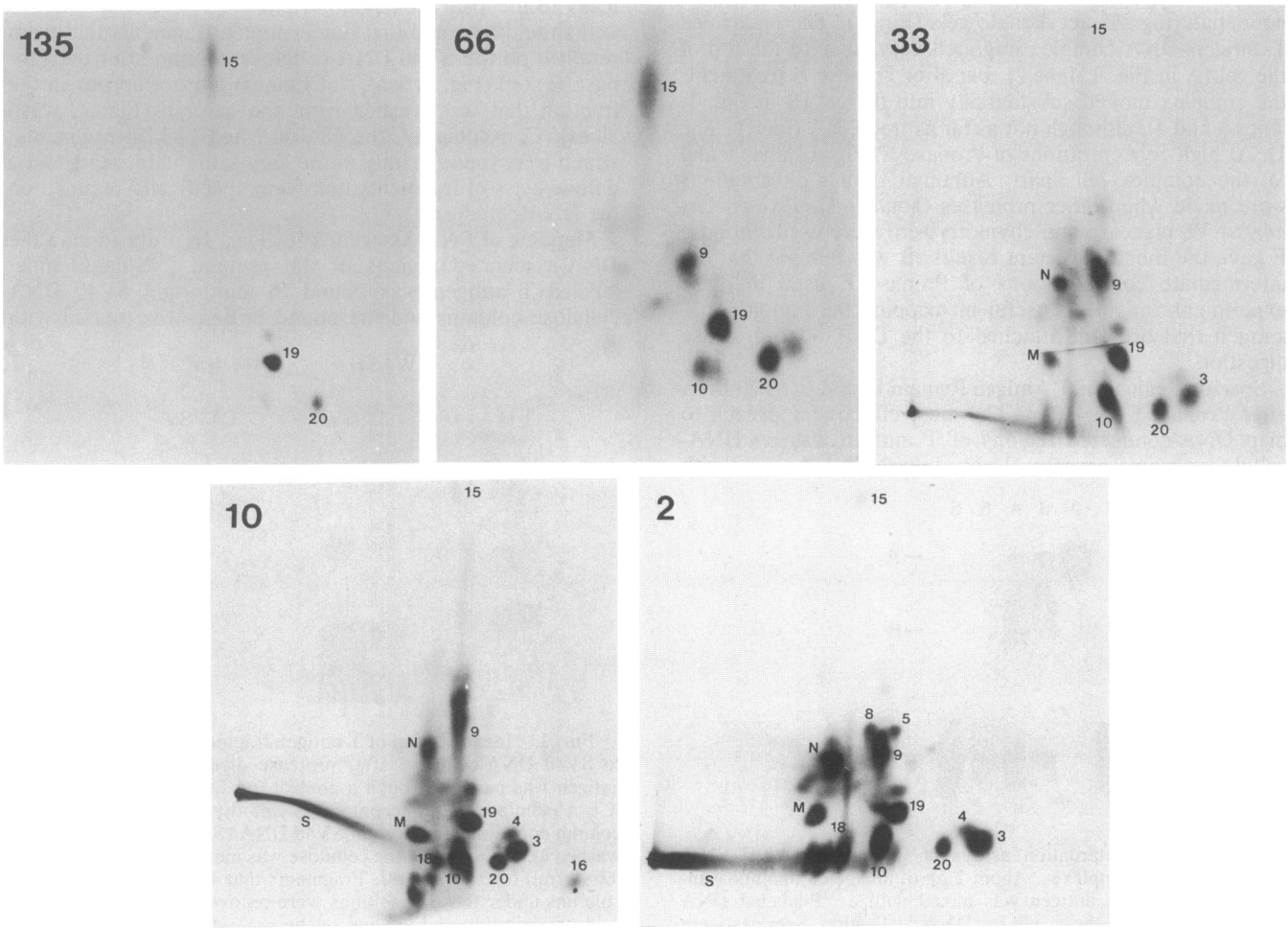

FIG. 3. Two-dimensional fingerprints of protected T-antigen fractions. [^{35}S]Labeled T antigen was applied to SV40 DNA-cellulose columns and incubated with various concentrations of Pronase E (135, 66, 33, 10, and 2 μg/ml, as indicated). After extensive washing, the bound fragments were eluted with buffer containing 1 M NaCl, precipitated with trichloroacetic acid, and exhaustively digested wih trypsin. The tryptic peptides were separated by two-dimensional electrophoresis and chromatography on thin-layer cellulose plates. Peptide spots are identified by number or letter (see Fig. 4 and ref. 8).

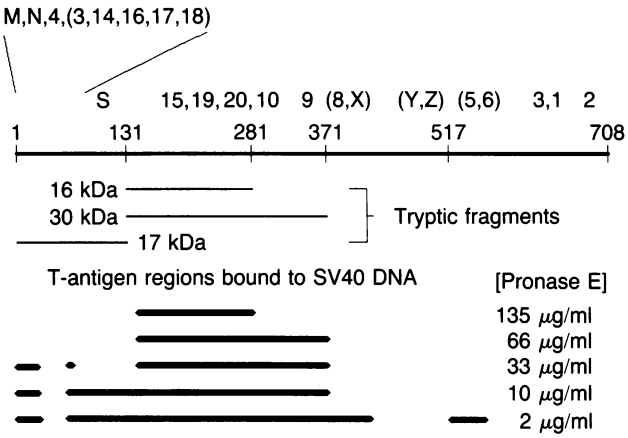

FIG. 4. Map of T-antigen regions bound to DNA. The linear maps shown at the bottom correspond to the regions identified by the fingerprints in Fig. 3. The top part of the figure depicts a linear map of the methionine-containing tryptic peptides of SV40 T antigen (8), as well as map positions for some partial tryptic fragments alluded to in the text.

and fingerprint pattern of the core, it is estimated that it extends approximately from residue 140 to 281, although the COOH-terminal end of this protected core could extend to residue 290. The last methionine-containing tryptic peptide in the core is peptide 20, which maps from residues 272 to 281; the next methionine is at residue 291. Since no additional methionine peptide is apparent in the fingerprint of the core region, the COOH end of this region lies between residues 281 and 290.

The resistant core described above probably defines the minimal domain that, under these conditions, makes a quasi-stable complex with SV40 DNA. In Fig. 1, it is apparent that the DNA–protein complex is unstable at high concentrations of Pronase E. At the approximate Pronase E concentrations used to isolate the core domain, a protein–DNA complex was only weakly detected in the gel retardation assay. In addition, previous mapping experiments (8) of the DNA-binding region, using partial tryptic fragments, showed that the DNA-binding activity of the 16-kDa fragment described above was very low. Furthermore, the yield of the core fragment has been consistently poor and this region has been consequently difficult to analyze. Nonetheless, the core region as defined here probably includes the critical components of the DNA-binding domain, which is entirely consistent with previous results from this laboratory (8) and those obtained by other groups (5–7, 9).

In contrast, the region of T antigen that remains bound to SV40 DNA in the presence of Pronase E at 66 μg/ml appears to make a much more stable complex with DNA. This region extends approximately to residue 371 (Fig. 4), although some fragments terminate between residues 304 and 358 (data not shown). At the concentrations of protease used to isolate this larger region, DNA–protein complexes are quite stable (Fig. 1). This observation, combined with previous studies of an analogous 30-kDa tryptic fragment (8), suggests that in the intact T-antigen molecule, additional sequences mapping between 281 and 371 are needed for the stable association of the protein with the DNA origin. Within this secondary region is a putative Zn^{2+} finger, a structure characteristic of many DNA-binding proteins. It is located between residues 302 and 320 and has the structure Cys-Xaa_2-Cys-Xaa_7-His-Xaa_3-His. Whether this sequence interacts with the DNA remains to be seen.

It is intriguing that the region of T antigen that is associated with the DNA in the presence of Pronase E at 33 μg/ml includes sequences from the NH_2-terminal region. Previous work from this laboratory (23) has shown that these sequences are unnecessary for binding to the SV40 replication origin (in fact, they may inhibit binding). It is unclear, however, whether they represent sequences that are protected from Pronase E digestion by the DNA or by other parts of the protein molecule. In any case, it is interesting that the pattern of bound peptides in this region is spotty. Residues 1–21 (peptides M and N) and 52–53 (peptide 3) are present, but not other regions in the first 131 residues. This is known because other methionine-containing tryptic peptides characteristic of this region are either missing or barely detectable. These include peptides 4, 14, 16, 17, 18, and S, all of which are easily identifiable in the fingerprints of intact T antigen or of a 17-kDa NH_2-terminal fragment (see Fig. 4 and ref. 8).

The NH_2-terminal region of T antigen is important for viral replication functions. Mutations in this region are lethal (24). Although these sequences are not needed for DNA binding *per se*, they are undoubtedly necessary for one or more event during DNA replication, and the protease-digestion experiments described here may have identified the portions of the molecule involved in this activity.

The role in DNA binding of other regions that remain bound to the DNA after treatment with lower concentrations of Pronase E is less clear. The region corresponding to peptide S (residues 68–130) contains several serines and threonines (residues 106–124) that are phosphorylated in the cell. Previous work in this laboratory (23), and by Mohr *et al.* (25), has demonstrated that these phosphate groups reduce the DNA-binding activity of T antigen. The small (factor of 1.5–2) effect detected in these studies may be due to a charge repulsion of the negatively charged phosphate groups on the protein and the DNA. This could be of functional importance or it may simply be due to the proximity of these phosphates to the DNA-binding domain, which probably starts around residue 140 (6). The significance of other bound regions (peptides 8 and 5) is not known, but these sequences are less likely to interact with the DNA. They may correspond to parts of the protein molecule that are partially protected due to the folding of the protein chain or to protein–protein interactions.

Several observations support the suggestion that the maps of DNA-bound fragments obtained in this study reflect the actual spatial relation between T antigen and DNA. First, the region that was present after digestion with the highest concentration of Pronase E correlates with the DNA-binding domain as mapped by other methods (5–9). Second, the next level of resistance corresponds with a region that has been shown to have an effect on DNA-binding activity (ref. 8 and this study) and that may have a DNA-binding Zn^{2+} finger. Third, other regions not directly involved in DNA binding are sensitive to digestion. This includes the nucleotide-binding domain/ATPase active site (residues 418–528) (26) and the COOH-terminal 80–100 amino acids, which have been shown to be dispensable for DNA-replication functions (24, 27). Other protease-sensitive regions, such as most of the sequences between residues 21 and 52 and others closer to the COOH-terminal end (Fig. 4), may be largely exposed in the T-antigen–DNA complex. The same regions may also be accessible to protease digestion in the native protein alone.

The approach described in this study may be useful in mapping the topography of other DNA–protein complexes.

I wish to acknowledge the helpful suggestions of Louis Levinger. This work was supported by Grant CA36618 from the National Cancer Institute.

1. Pabo, C. O. & Sauer, R. T. (1984) *Annu. Rev. Biochem.* 53, 293–321.

2. Berg, J. M. (1986) *Science* **232,** 485–487.
3. Miller, J., McLachlan, A. D. & Klug, A. (1985) *EMBO J.* **4,** 1609–1614.
4. Brown, R. S., Sander, C. & Argos, P. (1985) *FEBS Lett.* **186,** 271–274.
5. Clark, R., Peden, K., Pipas, J. M., Nathans, D. & Tjian, R. (1983) *Mol. Cell. Biol.* **3,** 220–228.
6. Paucha, E., Kalderon, D., Harvey, R. W. & Smith, A. E. (1986) *J. Virol.* **57,** 50–64.
7. Prives, C., Barnet, B., Scheller, A., Khoury, G. & Jay, G. (1982) *J. Virol.* **43,** 73–82.
8. Simmons, D. T. (1986) *J. Virol.* **57,** 776–785.
9. Strauss, M., Argani, P., Mohr, I. J. & Gluzman, Y. (1987) *J. Virol.* **61,** 3326–3330.
10. Tjian, R. (1978) *Cold Spring Harbor Symp. Quant. Biol.* **43,** 655–662.
11. Tjian, R. (1978) *Cell* **13,** 165–179.
12. DeLucia, A. L., Lewton, B. A., Tjian, R. & Tegtmeyer, P. (1983) *J. Virol.* **46,** 143–150.
13. Jones, K. A. & Tjian, R. (1984) *Cell* **36,** 155–162.
14. Deb, S., Tsui, S., Koff, A., DeLucia, A. L., Parsons, R. & Tegtmeyer, P. (1987) *J. Virol.* **61,** 2143–2149.
15. Simanis, V. & Lane, D. P. (1985) *Virology* **144,** 88–97.
16. Thummel, C., Tjian, R., Hu, S. & Grodzicker, T. (1983) *Cell* **33,** 455–464.
17. Levitsky, K., Chandrasekaran, K., Mora, T. & Simmons, D. T. (1983) *Int. J. Cancer* **32,** 597–602.
18. Stillman, B. W. & Gluzman, Y. (1985) *Mol. Cell. Biol.* **5,** 2051–2060.
19. Levinger, L. F. (1985) *J. Biol. Chem.* **260,** 14311–14318.
20. Simmons, D. T. (1984) *J. Biol. Chem.* **259,** 8633–8644.
21. Simmons, D. T., Martin, M. A., Mora, P. T. & Chang, C. (1980) *J. Virol.* **34,** 650–657.
22. Linke, H. K., Hunter, T. & Walter, G. (1979) *J. Virol.* **29,** 390–394.
23. Simmons, D. T., Chou, W. & Rodgers, K. (1986) *J. Virol.* **60,** 888–894.
24. Pipas, J. M., Peden, K. W. & Nathans, D. (1983) *Mol. Cell. Biol.* **3,** 203–213.
25. Mohr, I. J., Stillman, B. & Gluzman, Y. (1987) *EMBO J.* **6,** 153–160.
26. Bradley, M. K., Smith, T. F., Lathrop, R. H., Livingston, D. M. & Webster, T. A. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 4026–4030.
27. Tornow, J., Polvino-Bodnar, M., Santangelo, G. & Cole, C. N. (1985) *J. Virol.* **53,** 415–424.