Volume 13 Number 6 1985 Nucleic Acids Research

The TGGCA protein binds to the MMTV-LTR, the adenovirus origin of replication, and the BK virus enhancer

Joachim Nowock, Uwe Borgmeyer, Andreas W.Püschel, Ralph A.W.Rupp and Albrecht E.Sippel

Zentrum für Molekulare Biologie Heidelberg (ZMBH), Im Neuenheimer Feld 364, D-6900 Heidelberg, FRG

Received 1 February 1985; Accepted 18 February 1985

ABSTRACT

TGGCA-binding proteins are nuclear proteins with high affinity for double-stranded DNA homologous to the prototype recognition sequence 5'YTGGCANNNTGCCAR 3'. Their ubiquitous tissue distribution in higher vertebrates characterizes them as a class of highly conserved proteins which may exert a basic function. To obtain clues to this function, specific binding sites were mapped on three viral genomes. Recognition sites were identified in the enhancer region of the BK virus, in the LTR of the mouse mammary tumor virus, and in the origin of replication of adenovirus 12. The TGGCA-binding protein from HeLa cells appears to be identical to nuclear factor I described by others, which stimulates initiation of adenovirus DNA replication in vitro. However, data from MMTV, BKV, and from cellular genes suggest that this specific protein-DNA interaction may also be involved in the control of gene activity.

INTRODUCTION

Alteration of chromatin organization is the structural basis of the epigenetic code which controls the expression of genetic information during differentiation. Specific interactions of nonhistone proteins with recognition sites on DNA appear to be decisive regulatory elements in transcription and replication.

We have identified a DNA-binding protein with high affinity for specific sites in the flanking regions of the chicken lysozyme gene (1). DNAase I protection experiments revealed that the binding site contained a central region of diad symmetry, 5'TGGCANNNTGCCA3', which is recognized by the protein on double-stranded DNA (2). In reference to the half side of the recognition sequence, the protein was named 'TGGCA-binding protein'.

Since the binding sites are clustered around the

transcription unit of the lysozyme gene and since one coincides with a DNAase-hypersensitive site in the chromatin of the active gene (2,3), the specific TGGCA-binding protein-DNA interaction was thought to be involved in gene regulation. To specify its function, we have pursued two approaches: 1. an analysis of the tissue distribution and of the conservation across species boundaries of the binding protein, and 2. a search for TGGCA binding sites in other eucaryotic and viral DNA sequences.

The specific binding activity was not only found in many tissues (2), but appeared to be also conserved among higher vertebrates. Complementary to the widespread occurrence of the protein, several specific binding sites were identified on DNA from different sources. Binding sites on the genomes of mouse mammary tumor virus (MMTV), BK virus, and adenovirus were selected for a detailed analysis. This was based on the expectation that elucidating the function of the TGGCA protein-DNA interaction in these thoroughly studied viruses might cast light on its role in the host cell. An interesting outcome of these investigations was the functional identity of the TGGCA-binding protein with a recently identified, host-encoded protein (nuclear factor I) which enhances initiation of adenovirus DNA replication in vitro (4). However, comparing the data obtained from MMTV and BKV and from cellular genes (2,5), we propose that in eucaryotes the specific binding protein-DNA interaction is not tied sensu strictu to replication initiation but may serve in maintaining a differentiated state of chromatin structure.

MATERIALS AND METHODS

Recombinant plasmids

The following plasmids were used: pB2B3 (2); pB2H3, a subfragment of pB2B3 inserted into pBR322; pMMTV-LTR, a Pst I-Eco RI fragment of p2.6 (6) which contains the 3'LTR of MMTV and 222 bp of flanking mouse DNA cloned in pBR322; pBK (Dunlop), the Dunlop strain of BK virus (7) inserted into pBR322 at the Bam H1 site; pAd12-HindIII G (8).

Preparation of binding protein fractions

TGGCA-binding activity was obtained from nuclei by extrac-

tion with 300 mM NaCl and subsequent precipitation at 40% saturation of ammonium sulfate as described (2). These crude fractions were generally used for binding reactions. In one case the nuclear protein extract was further purified: nuclei from hen liver were extracted with buffer CB (10 mM Tris-HCl, 7.5, 1 mM EDTA, 7 mM mercaptoethanol, 10% glycerol, 0.5 mM PMSF, 2.5 KIU/ml aprotinin) containing 300 mM NaCl. The extract was passed over a DEAE-Sephacel column equilibrated with the same buffer. The flow-through was directly loaded onto a phosphocellulose column. The column was washed with two volumes of buffer CB containing 360 mM NaCl, and the TGGCA-binding activity was eluted in one step at 650 mM salt. After dialysis into buffer CB, 350 mM NaCL, the binding fraction was applied to a heparin-Sepharose column equilibrated with the same buffer. The column was washed with two volumes of buffer CB containing 420 mM NaCl and eluted with a step of 650 mM NaCl. Peak binding fractions were divided into small aliquots and stored at -70°C.

Nucleoids from E. coli strain DG 17 were prepared according to the procedure of Varshavsky (9) with slight modifications (W. Vielmetter, personal communication). Proteins were extracted from nucleoids with 300 mM NaCl as described for nuclei.

Filter binding assays and DNAase I protection experiments

Binding conditions, nitrocellulose filter assays and DNAase I protection experiments ('footprints') followed our published procedures (2). The state of purification of the extracts was sufficient to demonstrate only one kinetically homogeneous population of binding proteins. On additional purification apparent binding constants did not change.

RESULTS

Occurrence of the TGGCA-binding protein in different species

Although TGGCA-binding was initially detected in a highly specialized organ, the chicken oviduct, in connection with the lysozyme gene, we soon realized that homologous binding activity could be demonstated in nuclei of all analyzed chicken tissues (2). To test whether it is conserved beyond species boundaries, we performed binding studies with nuclear protein extracts from species of divergent classes of the phylogenetic tree. Nuclear

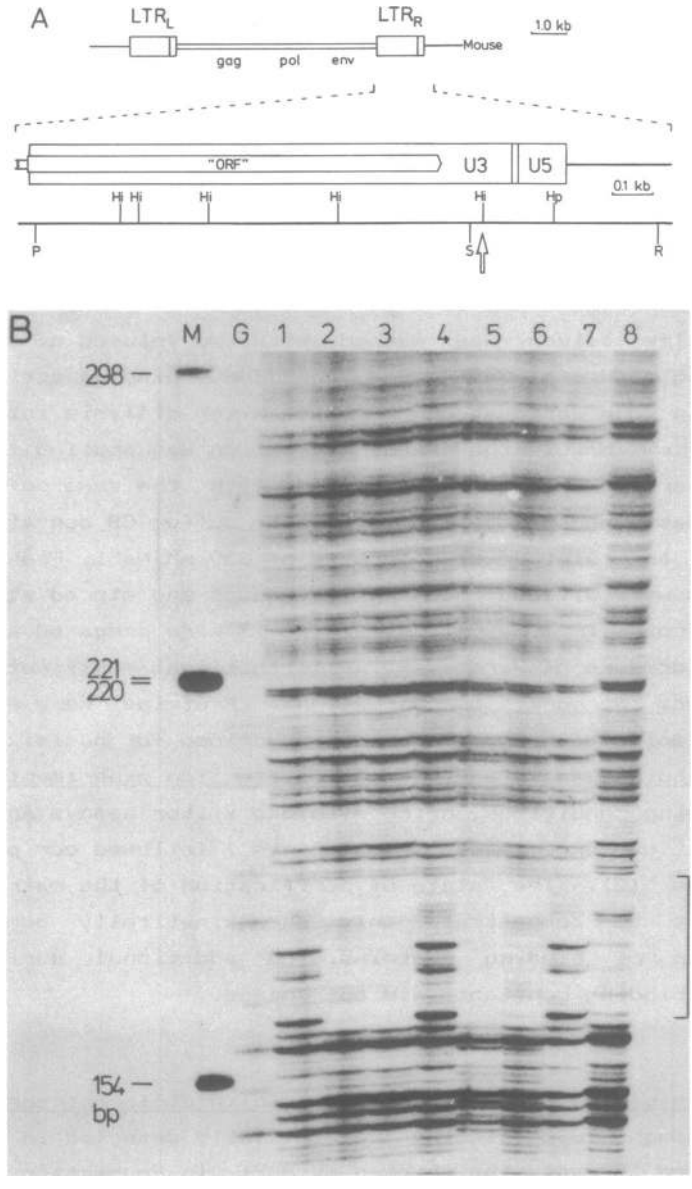

Fig. 1. TGGCA protein binding to MMTV-LTR.
A, Map of integrated proviral DNA. LTR L and LTR R indicate the 5' and 3' long terminal repeats. gag, pol, env designate the coding regions of the core protein, the reverse transcriptase, and the envelope protein, respectively. LTR R is shown in enlarged scale with restriction sites for Eco RI (R), Pst I (P), Sac I (S), Hpa II (Hp) and Hinf 1 (Hi). "orf" open reading

frame. The location of the binding site for the TGGCA-binding protein is indicated by an arrow.
B, DNAase I protection experiment. Three fmole of the Pst I-Hpa II fragment of pMMTV-LTR, labeled at the Hpa II site, and 100 ng Eco RI-digested pBR322 as unspecific competitor were incubated for 20 min at 25°C with various dilutions of nuclear extracts from chicken oviduct or mouse liver. Final extract dilutions were 1/500 (lane 2) and 1/300 (lane 3) for oviduct, 1/200 (lane 5), 1/300 (lane 6) and 1/500 (lane 8) for mouse liver (2). Incubations with BSA (lanes 1, 4, 7) served as control. After equilibration of the binding reaction, 60 ng DNAase I were added and incubated for 3 min. Endonuclease digested DNA was separated on a sequencing gel together with a size marker (M) and a G-specific sequencing reaction (G) (33).

extracts from mouse liver (2) and human Hela (Fig. 3) and KB cells contained a DNA-binding activity specific for the TGGCA motive. Extracts from Drosophila Kc cells (10) and from E. coli nucleoids showed only nonspecific binding (data not shown). From these examples we conclude that at least in higher vertebrates a class of proteins exists which contains a highly conserved DNA-binding domain for the recognition of the TGGCA motive.

Binding sites on viral DNA recognized by the TGGCA protein

The widespread occurrence of the TGGCA-binding protein led us to expect, by dialectic reasoning, a corresponding frequency of the respective recognition sequence on DNA. We therefore searched DNA data banks for sequences with at least 70% homology to the consensus sequence 5'TGGCANNNTGCCA 3'. Selected DNAs containing homologous sequences were tested directly for specific binding with nuclear extract from chicken oviduct by nitrocellulose filter binding assays. From the positives we finally analyzed sites on three viral genomes.

Mouse mammary tumor virus We chose mouse mammary tumor virus because transcription of the proviral DNA is positively regulated by glucocorticoid hormone (11,12), a situation comparable to the steroid hormone control of lysozyme gene expression in the chicken oviduct. Initial mapping experiments located a specific binding site on identical restriction fragments of both long terminal repeats (LTR) (data not shown). Fig. 1 shows a footprint (13), using a subfragment of the 3'LTR. A protein-protected region extends from position -80 to -58 with respect to the transcriptional start site. It has homology to the consensus recognition sequence (Fig. 4). The TGGCA binding

site is adjacent to the promoter proximal binding site for the glucocorticoid receptor (14). Nuclear extracts from hen oviduct or mouse liver yielded identical footprints. The presence of a recognition site for the TGGCA-binding protein on the LTR of MMTV appears to be unique among retroviruses because homology screening of all LTR sequences available in the EMBL sequence library (July 1984) and partial testing by nitrocellulose filter binding did not reveal another binding site.

BK virus The human papovavirus BKV (Fig. 2A) has a genetic complement and a structural organization similar to the monkey virus SV40 and the mouse polyoma virus (15). Binding experiments with HindIII-digested DNA of BKV (Dunlop), a strain very similar to prototype BKV, showed that on incubation with nuclear extract of chicken liver or HeLa cells the HindIII C fragment was specifically retained on nitrocellulose filters (data not shown). This fragment contains the origin of replication and a triplicate tandem repeat on the 'late' side. Footprint analysis showed two sites protected from DNAase I digestion which mapped at homologous positions on each of the outer members of the repeat triplet (Fig. 2B). This was confirmed by individual mapping on separate fragments (data not shown). The protein-protected regions coincide with a sequence homologous to the consensus sequence (Fig. 4). Interestingly, the central repeat did not show any binding. In this repeat unit a deletion of 18bp alters one side of the recognition sequence (see Fig. 4), thereby eliminating binding.

SV40 does not contain a recognition site for the TGGCA-binding protein (our unpublished observation). Although considerable sequence homology to BKV exists, the regions of the tandem repeats are dissimilar. It has been suggested that the repeated sequences did not diverge from a common viral ancestor but have been acquired independently of the host genome (15).

Adenovirus 12 Binding studies with adenovirus 2 DNA demonstrated binding of both terminal restriction fragments (data not shown) indicating that sequences in the inverted terminal repeats were involved in this protein-DNA interaction. Homologous binding to the left-hand terminal Hga I restriction fragment of adenovirus 12 was found by footprint analysis. The

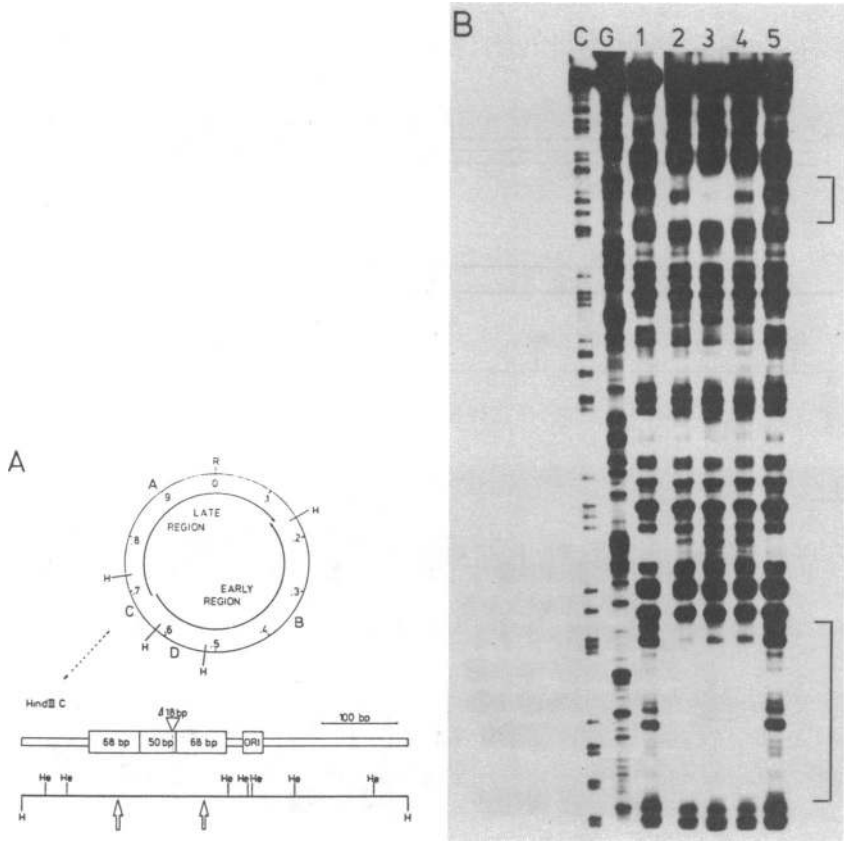

Fig. 2. TGGCA protein binding to BKV
A, Map of the BK virus. Restriction sites for Eco RI (R) and Hind III (H) are shown on the map of the viral genome. The restriction fragment Hind III C is drawn at enlarged scale showing the origin of replication and the three tandem repeats. Restriction sites for Hind III (H) and Hae III (He) are given. The arrows indicate the position of the binding sites for the TGGCA protein.
B, DNAase I protection experiment. Bam H1 linkers were ligated to the 216bp Hae III subfragment containing the tandem repeats and cloned in the plasmid pA II (7). The fragment was recovered by excision at unique restriction sites in the plasmid poly-linker and was labeled by 'Klenow' fill-in reaction at the origin proximal side. 4.5 fmoles of fragment were incubated with dilutions of a purified fraction of TGGCA-binding protein from chicken liver (see Methods). All other conditions were as described in legend to Fig. 1. Incubations were performed with 1/20 (lane 2), 1/50 (lane 3) and 1/100 diluted binding protein fraction (lane 4) or with BSA (lanes 1 and 5). G,C: G- and C-specific sequencing reactions.

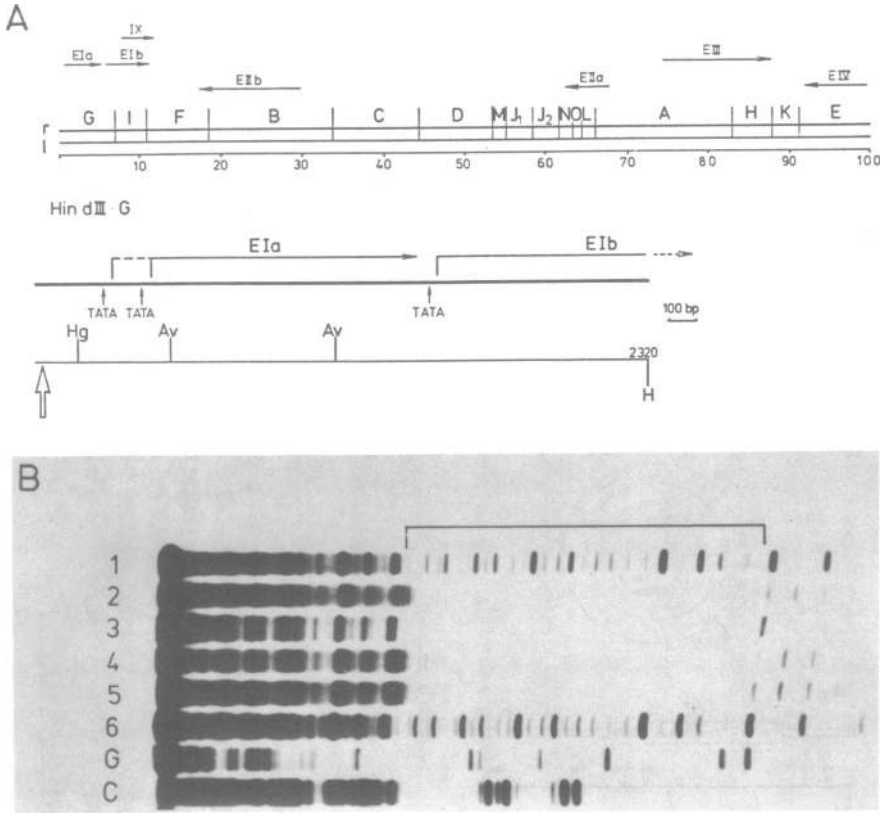

Fig. 3. TGGCA protein binding to the left-hand terminus of adenovirus 12.
A, Genomic map of adenovirus 12 showing the cleavage sites for Hind III. Restriction fragments are designated in alphabetical order according to size. Horizontal arrows indicate early transcription units. Hind III fragment G containing the left-hand terminus is shown at enlarged scale. TATA denotes the position of the TATA boxes related to the early promoters E1a and E1b. The large vertical arrow indicates the recognition site of the TGGCA-binding protein. Restriction sites for Hga I (Hg) and Ava II (Av) are given.
B, DNAase I protection experiment. The Eco RI-Hga I fragment was labeled with [^{32}P] at the 3'-end of the Eco RI site (the Eco RI site is derived from a linker ligated to the natural terminus of fragment Hind III G to generate plasmid pAd12-HindIII G). Five fmoles of fragment were incubated with dilutions of nuclear extracts from HeLa cells (lanes 2 and 5), chicken oviduct (lane 3), mouse liver (lane 4), or with BSA (lanes 1, 6) and were processed further as described in legend to Fig. 1. G, C: G- and C-specific chemical sequencing reactions.

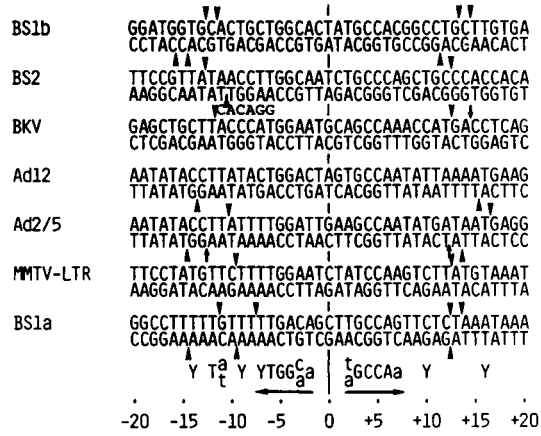

Fig. 4. Compilation of DNA sequences containing regions protected from DNAase I digestion by TGGCA-binding protein and nuclear factor I.
Sequences were aligned at the central homology. Vertical arrows delineate the boundaries of the DNAase protected domain on the respective DNA strand. Arrows of different type represent protection at different sites with identical sequence. When assignment of a boundary was ambiguous, an interval is shown by two symbols of the same type. Conserved nucleotides within the protected region are indicated. A central core of twofold rotational symmetry is underlined. The characters above the BKV sequence show the base alterations of the middle repeat caused by a 18bp deletion. References for DNAase protection and/or DNA sequence data: BS1a, BS1b, BS2 (2); BKV (15); Ad12 (16); Ad5 (4); Ad2 (18); MMTV-LTR (32, N. Kennedy, personal communication).

sequences of the inverted terminal repeats are highly conserved between adenovirus 2 and 12 (16). A region located at position 18-46 was protected from DNAase cutting by a protein in nuclear extracts from chicken oviduct, mouse liver, and HeLa cells (Fig. 3B). The protected DNA sequence contains the TGGCA consensus motive (Fig. 4). It is present in many adenovirus serotypes (4), and is embedded in a region which constitutes the origin of replication (16). This stresses the importance of this protein-DNA interaction for the control of the viral replication cycle. Indeed, a protein from uninfected HeLa cells has been described which generates a DNAase I protection pattern identical to that shown here on the homologous sequence of adenovirus serotype 5 (17) and serotype 2 (18). This protein, termed nu-

clear factor I, stimulates initiation of adenovirus DNA replication in vitro (17). We conclude that the TGGCA-binding proteins we detected in various species and the nuclear factor I contain a DNA-binding domain with identical sequence specificity. They most likely represent a class of homologous proteins (see Discussion).

Comparison of the protein-protected DNA sequences

Fig. 4 compiles the DNA sequences protected from DNAase I digestion by binding of the TGGCA protein to the three viral recognition sites. In addition, footprint data of the binding sites on the 5' flanking region of the lysozyme gene (2) and of adenovirus serotype 5 (4) and 2 (18) are included. When the DNA sequences are aligned at the symmetry axis of the palindromic homology region, the coincidence of DNA sequence and DNAase I protected domain becomes apparent. Slight variations at the boundaries of the protected areas may derive from imprecise assignment due to a locally low cutting frequency of the endonuclease.

A consensus sequence can be derived from the binding sites: 5'YTGG$\frac{c}{a}$a-*-$\frac{t}{a}$GCCAa 3' (Fig. 4). Although consistent with the consensus sequence 5'TGGCANNNTGCCA 3' initially deduced from the binding sites BS1a, BS1b and BS2 in the lysozyme gene region (2), the additional data allow a more precise definition. It appears that the consensus sequence can be extended by one base pair on either side. Further, at position -3 an A will be tolerated for a C and, for symmetry reasons, a T at position +3 for a G (Fig. 4). The central bases of the consensus sequence (position -2/+2) conform less stringently to symmetry and may be of lesser importance for interaction with the protein. It is not clear whether the homologies outside the symmetry are used for recognition.

Inspecting the regions adjacent to the consensus sequence, an additional feature becomes apparent. In a subgroup of binding sites (adenovirus 2, 5, and 12, MMTV-LTR, BS1a), the symmetry region is flanked by A/T-rich stretches at homologous positions. Whether these sequences have any bearing on the effect of the protein-DNA interaction is presently unknown.

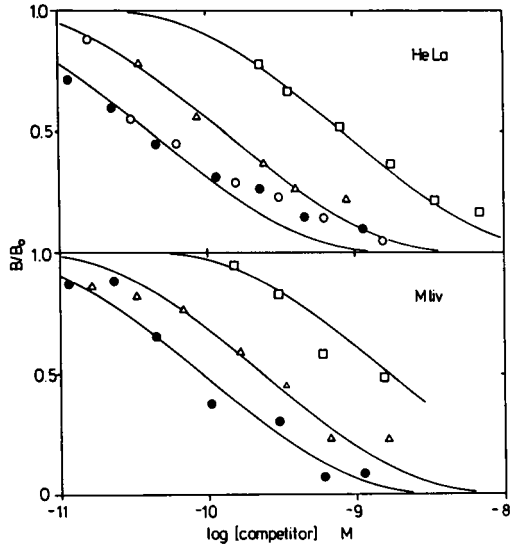

Fig. 5. Equilibrium competition binding.
Binding reactions contained one fmole Eco RI-linearized [^{32}P] pAd12-HindIII G as labeled ligand and increasing amounts of Eco RI-linearized DNA of pAd12-HindIII G (●), pB2H3 (○), pMMTV-LTR (△) and pBR322 (□) as unlabeled competitors in 100 ul assay volume. Concentrations of nuclear extracts from HeLa cells and mouse liver (Mliv) were set to yield 50% binding of input [^{32}P] DNA in the uncompeted reaction (Bo). B/Bo is labeled, filter-bound DNA in the presence of the indicated amount of competitor as fraction of uncompeted binding.

Relative affinities of TGGCA proteins from various species to different DNA binding sites

The variability of certain bases within the homology region could reflect a species-specific binding preference. We therefore performed competition binding experiments to compare relative affinities of linearized plasmids pB2H3, pAd12-HindIII G, and pMMTV-LTR to TGGCA-binding proteins from chicken oviduct, HeLa cells, and mouse liver. Fig. 5 shows that pB2H3 and pAd12-HindIII G are equally bound by nuclear extract from HeLa cells. Affinity to pMMTV-LTR is three- to fourfold lower. The plasmid pB2H3 contains the two closely spaced binding sites BS1a and BS1b of the lysozyme gene region (1). Since BS1a has only one-tenth the affinity of BS1b, its contribution to the overall binding strength of the B2H3 fragment is insignificant (2).

Identical relative binding affinities are observed when competition experiments were conducted with nuclear extracts from mouse liver (Fig. 5) or hen oviduct (data not shown). Thus, the DNA recognition properties of the TGGCA-binding proteins from man, mouse and chicken are highly conserved and do not display a species- or tissue-specific variation.

DISCUSSION

A recognition sequence of twofold rotational symmetry

Comparison of binding sites on viral and eucaryotic DNA with high affinity for a nuclear protein revealed a conserved core sequence of twofold rotational symmetry from which the idealized prototype 5'YTGGCANNNTGCCAR 3' can be deduced. We think that it is mainly this region of dyad symmetry which is specifically recognized by the DNA-binding protein (see also ref. 2). Only limited base alterations are found in the natural binding sites when compared with the ideal prototype. The highest stringency seems to be imposed on the triplet 5'TGG 3' on each half-site. The base pair C/G at position -3/+3 (see Fig. 4) may be exchanged on one half-site by A/T without detectable loss of binding strength (compare pB2H3 with pAd12-HindIII G in Fig. 5). A double-sided replacement as in MMTV-LTR (see Fig. 4) leads to a three- to fourfold reduction of affinity (Fig. 5). As a general conclusion it can be stated that a balanced set of a few base alterations outside the 5'TGG 3' triplets generates a class of binding sites with comparable affinity. Remarkably, binding sites which completely match the ideal prototype have not been found so far. Thus, either optimal protein-DNA interaction has an inherent asymmetry or increased symmetry results in such a high binding affinity that it would be physiologically detrimental. We are currently testing these possibilities with a synthetic DNA fragment containing the 'ideal' recognition site.

The dyad symmetry of the consensus sequence led us to propose that the binding site is occupied by a protein dimer of the same symmetry (2). Since the highly conserved triplets 5'TGG 3' are 3.4 nm apart on DNA in B conformation, the DNA-binding domains of the protein monomers would be positioned on one face of the DNA helix, probably at adjacent major grooves.

TGGCA-binding proteins/nuclear factor I constitute a class of highly conserved proteins

DNA-binding proteins with recognition specificity for the consensus sequence described above have been detected in several vertebrates (2,4,18, this paper). For one species, a general occurrence in many tissues has been demonstrated (2). Several lines of evidence argue that these proteins, termed TGGCA-binding proteins or, in the case of HeLa cells, nuclear factor I, perform identical functions and are similar in structure: 1. They bind - per definition - selectively and with high affinity to homologues of the recognition sequence 5'YTGGCANNNTGCCAR 3'. Different variants of this sequence are bound with the same relative affinities by TGGCA-binding proteins from different sources. Absolute affinities for a particular DNA binding site are also comparable: an apparent equilibrium dissociation constant of $K_D = 2 \times 10^{-11}$ M at 150 mM NaCl has been determined for factor I from HeLa cells (18), TGGCA-binding protein from hen oviduct yielded a $K_D = 6 \times 10^{-12}$ M at 100 mM NaCl (1). 2. The only functional assay established so far besides specific DNA binding is an in vitro system which replicates template DNA containing the origin of replication of adenovirus (19-21). It has been shown by this assay that nuclear factor I enhances the initiation reaction (4). A protein fraction from mouse, hamster (18), and chicken (P. Leegwater, P. van der Vliet, R. Rupp, J. Nowock, and A. E. Sippel; unpublished observation) with specificity for the TGGCA binding site can be substituted for HeLa factor I. 3. The various proteins protect in vitro a remarkably constant 23 to 25 base pair region of DNA from DNAase digestion. 4. The sedimentation coefficient has been estimated for both the HeLa factor I (4) and the chicken TGGCA-binding protein (R. Rupp, unpublished result) to be 3.4 S.

Functional aspects

We have detected and analyzed binding sites on viral DNA which are recognized by a host-encoded DNA-binding protein. The effects of this specific protein-DNA interaction on the metabolism of well studied viruses may provide clues to its role in the host. A prerequisite for functional considerations is that the binding sites determined by in vitro techniques are of in

vivo significance for the virus. Several in vitro mutational approaches in combination with DNA transfection experiments had not yet defined the TGGCA binding site on the MMTV-LTR as an independent control element (6,22-24). We have constructed a mutant by a 3 base pair insertion into the center of dyad symmetry of the recognition site, which eliminates binding by the TGGCA-binding protein. After linking the mutant LTR to a thymidine kinase gene and transfection into mouse Ltk- cells, the tk+ transformants showed a strongly reduced glucocorticoid hormone response compared with wild type (J. Nowock, H. Ponta, U. Borgmeyer, M. Theisen, and A. E. Sippel; manuscript in preparation). This result establishes the functional significance of this binding site. The importance of the binding site in BKV is inferred from its conservation in the tandem repeats of another human polyomavirus, the JC virus (25). Except for the TGGCA binding motive, the two papovaviruses share little further sequence homology in their repeats. Finally, the involvement of the binding sequence on the inverted terminal repeats of adenovirus in DNA replication has already been mentioned above.

At present it is unknown whether the role of TGGCA-binding protein/nuclear factor I in initiation of adenovirus DNA replication is homologous to a function of the host or whether it represents an assimilation of this protein into heterologous viral functions. The inverted terminal repeats of human adenovirus subgroups D and E and of many nonhuman adenoviruses lack the TGGCA motive. This sequence may either have been lost or may not have been acquired during evolution. At least it appears to be dispensible for replication. That initiation enhancement of replication is the sole function in adenovirus is put into question by our finding of other TGGCA binding sites in adenovirus 2 located internally (J. Nowock, unpublished).

The intriguing association of TGGCA binding sites with elements defined as transcriptional enhancers in BKV (7) and MMTV (22) suggested that interaction of the TGGCA-binding protein with the recognition sequence could be a component of the molecular complex of certain activator mechanisms. Using the chloramphenicol acetyltransferase (CAT) assay to evaluate enhancer activity (26), small fragments containing two different

TGGCA recognition sites of the lysozyme gene region failed to induce CAT activity (M. Theisen, unpublished results). The close proximity of a minor glucocorticoid receptor binding site (14) and the TGGCA recognition site on the MMTV-LTR made experiments with these sequences not feasible. On the other hand, the major glucocorticoid receptor binding region on the LTR, when physically separated from the TGGCA binding site, still acted - though moderately - as a hormone-responsive enhancer element (22). It thus appears that the specific TGGCA-binding protein-DNA interaction does not by itself lead to an enhancement of transcription, although it can act as an auxiliary element in this process.

Clues which may elucidate the role of the specific TGGCA protein-DNA interaction come from an alternative experimental approach which addresses the structural organization of genes in chromatin. The first activation of the chicken lysozyme gene in oviduct (27,28) and of the integrated MMTV in the mouse mammary gland (29) requires a process of differentiation. The early-to-late transcriptional switch of viruses represents a similar discrete developmental transition. Differential gene activation is accompanied by alterations of chromatin structure. Especially short regions of perturbed nucleosomal array, which are particularly open to trans-acting factors (and endonucleases, hence the operational term DNAase-hypersensitive sites), are regarded as critical elements of gene regulation. An interesting coincidence of individual hypersensitive sites with TGGCA recognition sequences can be found in the lysozyme gene region of hen oviduct (2,3,30), in the 5' region of the myc gene of lymphoblastoid cells and Burkitt lymphomas (5), and in the LTR of MMTV derivatives stably integrated into mouse L cells (31). Similar data for BKV are not available. When the dynamics of these hypersensitive sites associated with TGGCA recognition sites were analyzed, a strict correlation was observed between their appearance and the expression of the respective gene. However, the omnipresence of the TGGCA-binding protein makes it unlikely that the protein alone can establish a hypersensitive site or that it can act as a switch changing the state of gene expression. It might rather exert a more conservative function

in conjunction with other elements of a multifactorial control unit, which determines tissue- and stage-specific expression of a gene.

ACKNOWLEDGEMENTS

We thank P. Gruss for providing plasmid pBK (Dunlop), B. Groner for p2.6 and S. Stabel for pAd12-HindIII G. M. Theisen of our research group communicated unpublished results. L. Hennighausen brought to our attention the TGGCA recognition sequence on BK virus DNA. This work was supported by a grant from the Bundesministerium für Forschung und Technologie (BCT 0364/1).

REFERENCES

1. Nowock, J. and Sippel. A.E. (1982) Cell 30, 607-615.
2. Borgmeyer, U., Nowock, J., and Sippel, A.E. (1984) Nucleic Acids Res. 12, 4295-4311.
3. Fritton, H., Sippel, A.E., and Igo-Kemenes, T. (1983) Nucleic Acids Res. 11, 3467-3485.
4. Nagata, K., Guggenheimer, R.A., Enomoto, T., Lichy, J.H., and Hurwitz, J. (1982) Proc. Natl. Acad. Sci. USA 79, 6438-6442.
5. Siebenlist, U., Hennighausen, L., Battey, J., and Leder, P. (1984) Cell 37, 381-391.
6. Hynes, N., van Ooyen, A.J.J., Kennedy, N., Herrlich, P., Ponta, H., and Groner, B. (1983) Proc. Natl. Acad. Sci. USA 80, 3637-3641.
7. Rosenthal, N., Kress, M., Gruss, P., and Khoury, G. (1983) Science 222, 749-755.
8. Schrier, P.I., Bernards, R., Vaessen, R.T.M.J., Houweling, A., and van der Eb, A,J. (1983) Nature 305, 771-775.
9. Varshavsky, A.J., Nedospasov, S.A., Bakayev, V.V., Bakayeva, T.G., and Georgiev, G.P. (1977) Nucleic Acids Res. 4, 2725-2745.
10. Echallier, G. and Ohanessian, A. (1970) In vitro 6, 162-172.
11. Ringold, G.M., Yamamoto, K.R., Tomkins, G.M., Bishop, J.M., and Varmus, H.E. (1975) Cell 6, 299-305.
12. Young, H.A., Scolnik, E.M., and Parks, W.P. (1975) J. Biol. Chem. 250, 3337-3343.
13. Galas, D.J. and Schmitz, A. (1978) Nucleic Acids Res. 5, 3157-3170.
14. Scheidereit, C., Geisse, S., Westphal, H.W., and Beato, M. (1983) Nature 304, 749-752.
15. Seif, I., Khoury, G., and Dhar, R. (1979) Cell 18, 963-977.
16. Stillman, B.W., Topp, W.C., and Engler, J.A. (1982) J. Virology 44, 530-537.
17. Nagata, K., Guggenheimer, R.A., and Hurwitz, J. (1983) Proc. Natl. Acad. Sci. USA 80, 6177-6181.
18. Rawlins, D.R., Rosenfeld, P.J., Wides, R.J., Challberg, M.D., and Kelly, T.J., Jr. (1984) Cell 37, 309-319.

19. Challberg, M.D. and Kelly, T.J., Jr. (1979) Proc. Natl. Acad. Sci. USA 76, 655-659.
20. Tamanoi, F. and Stillman, B.W. (1982) Proc. Natl. Acad. Sci. USA 79, 2221-2225.
21. van Bergen, B.G.M., van der Ley, P.A., van Driel, W., van Mansfeld, A.D.M., and van der Vliet, P.C. (1983) Nucleic Acids Res. 11, 1975-1989.
22. Chandler, V.L., Maler, B.A., and Yamamoto, K.R. (1983) Cell 33, 489-499.
23. Majors, J. and Varmus, H.E. (1983) Proc. Natl. Acad. Sci. USA 80, 5866-5870.
24. Lee, F., Hall, C.V., Ringold, G.M., Dobson, D.E., Luh, J., and Jacob, P.E. (1984) Nucleic Acids Res. 12,
25. Frisque, R.J., Bream, G.L., and Cannella, M.T. (1984) J. Virology 51, 458-469.
26. Gorman, C.M., Moffat, L.F., and Howard, B.H. (1982) Mol. Cell. Biol. 2, 1044-1051.
27. Kohler, P.O., Grimley, P.M., and O'Malley, B.W. (1969) J. Cell Biol. 40, 8-27.
28. Oka, T. and Schimke, R.T. (1969) J. Cell Biol. 41, 816-831.
29. Bentvelzen, P. (1972) Int. Rev. Exp. Pathol. 11, 259-297.
30. Sippel, A.E., Nowock, J., Theisen, M., Borgmeyer, U., Strech-Jurk, U., Bonifer, C., Igo-Kemenes, T., and Fritton, H.P. (1985) in Co-ordinated Regulation of Gene Expression, Clayton, R.M. and Truman, D.E.S. Eds., Plenum Press, New York, London, in press.
31. Zaret, K.S. and Yamamoto, K.R. (1984) Cell 38, 29-38.
32. Kennedy, N., Knedlitschek, G., Groner, B., Hynes, N., Herrlich, P., Michalides, R., and van Ooyen, A.J.J. (1982) Nature 295, 622-624.
33. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.