Volume 15 Number 23 1987 Nucleic Acids Research

Chicken liver TGGCA protein purified by preparative mobility shift electrophoresis (PMSE) shows a 36.8 to 29.8 kd microheterogeneity

Ralph A.W.Rupp and Albrecht E.Sippel

Zentrum für Molekulare Biologie der Universität Heidelberg (ZMBH), Im Neuenheimerfeld 282, D-6900 Heidelberg, FRG

Received September 17, 1987; Revised and Accepted October 30, 1987

ABSTRACT

The TGGCA protein, the chicken homologue of HeLa cell NF-I, was purified to homogeneity from liver tissue by a procedure which includes preparative mobility shift electrophoresis (PMSE) as the final step. PMSE was here adjusted for the isolation of the TGGCA protein, but can be used as a general method to characterize the protein moiety of specific DNA-binding proteins. The TGGCA protein is a family of 6 protein species, which show minor differences in molecular weight from 36.8kd to 29.8kd. This microheterogeneity differs from the size distribution reported for HeLa cell NF-I polypeptides. All species of the TGGCA protein bind identically to a synthetic DNA-binding site and appear to be highly related in primary structure. We discuss the possible functional importance of this microheterogeneity.

INTRODUCTION

The TGGCA protein was originally detected by us in nuclear extracts from chicken oviduct tissue (1). We soon realized that this specific DNA-binding activity was not tissue-specific but could be found in comparable amounts in a variety of chicken cells (2). Nagata et al. (3) characterized independently NF-I, a DNA-binding protein from HeLa cells with identical sequence-specificity to the chicken TGGCA protein (4). Both proteins are most likely members of a highly conserved protein family, since identical DNA-binding activities were ubiquituously found in all tested eucaryotic species (3, 5-7). They are therefore candidates for basal functions of the eucaryotic cell.

The binding of NF-I to its binding site within the Adenovirus type 2 inverted terminal repeat enhances the initiation reaction of viral replication **in vitro** (5, 8) and **in vivo** (9). We have shown that the TGGCA protein from chicken liver can

substitute for NF-I in the reconstituted system **in vitro** (4). The involvement of NF-I/TGGCA protein in the cellular replication process has still to be proven. However, there is strong evidence for a function in transcriptional regulation. TGGCA protein/NF-I acts as a transcription factor on viral and cellular enhancer/promoter elements (6, 10-14). Purified NF-I (15) turned out to be identical with CCAAT-binding transcription factor CTF and stimulates **in vitro** transcription from the human α-globin promoter (16).

Here we report the purification of the TGGCA protein from chicken liver. It is identified as a family of 6 protein species by a combination of preparative mobility shift electrophoresis (PMSE) and SDS-PAGE. TGGCA proteins show a microheterogeneity in molecular weight and share a common peptide pattern on S.aureus V8-proteolysis. While PMSE might generally be applied for the purification of trans-acting factors, we demonstrate here that in combination with SDS-PAGE the method can be conveniently used to identify the molecular weights of DNA-binding proteins by gel electrophoretic procedures as conventionally used for analytical methods.

MATERIALS AND METHODS

Materials were purchased from: DNaseI - Cooper biomedical; Staphylococcus aureus V8-protease - Boehringer (Mannheim); phosphocellulose - Whatman; DEAE-Sephacel, Heparin-Sepharose - Pharmacia; ds calf-thymus DNA-cellulose - Sigma.

Prepurification of TGGCA protein

The purification started with 500g liver tissue from 6 week old chicken. All steps were carried out at 4°C. Preparation of the nuclear extract and the buffer systems used were described previously (2). The following modifications were introduced for large scale preparations: A polytron-homogenizer was used for tissue disruption and $(NH_4)_2SO_4$-precipitation was left out. Crude nuclei were washed twice with buffers HB(0.5% Triton X100) and SB, respectively, and preeluted with buffer PB(100mM). After elution with 0.3M NaCl, nuclei were pelleted (14500xg, 15min). NP40 was added to the supernatant (0.025%(w/v) final concentration), which was then diluted to

0.25M NaCl. The nuclear extract (NE,310ml) was clarified at 158.000xg for 30 min in a 45Ti-rotor and incubated for 15 min with DEAE-Sephacel (120ml), equilibrated in column buffer CB250 (10mM TrisHCl, pH7.5, 1mM EDTA, 10% glycerol, 0.025% NP40, 0.5mM PMSF, 2.5KIU/ml Aprotinin, 7mM ß-mercaptoethanol and 250mM NaCl). The suspension was filled into a wide column (5x55cm) and the total protein flow-through was collected (DEAEFT, 550ml). DEAEFT was applied to a phospho-cellulose column (40ml), equilibrated in CB250. The column was washed and eluted with a linear salt gradient from 250-1000mM NaCl. TGGCA-binding activity was detected in fractions from 510-600mM NaCl. The activity peak fractions were pooled (PCF, 90ml) and dialysed into CB350. The dialysed PCF was loaded on a Heparin-Sepharose column (12ml), equilibrated in CB350. The TGGCA binding activity eluted between 560 and 650mM NaCl from a linear salt gradient (350-1000mM NaCl). After dialysis into CB150, HSF (22ml) was finally chromatographed on double-stranded calf-thymus DNA-cellulose (4ml, equilibrated in CB150). Here the TGGCA-binding activity was found in fractions from 330-370mM NaCl of a linear salt gradient (150-1000mM NaCl). Pooled fractions (DCF, 3.5ml) were dialysed into CB100. Aliquots were frozen in liquid nitrogen and stored at -70°C.

Column fractions were screened for TGGCA-binding activity by analytical mobility shift assays (17, 18). Protein concentrations were determined with the Biorad Protein assay kit using bovine serum albumin as standard.

Preparative mobility shift electrophoresis

For all DNA-binding reactions we used sBS2, a 100bp DNA fragment containing two consensus recognition sequences for the TGGCA protein (6, 13). The PMSE-binding reactions were performed in 25µl binding buffer BB (10mM HEPES, pH8.0, 100mM NaCl, 2mM DTE, 0.1mM EDTA) including 500ng poly(dI-dC).(dI-dC). Careful titration of the DC-fraction showed that 4pmol DNA-binding sites ($2.9x10^5$cpm/µg) and a 2-fold excess of TGGCA-binding activity were needed to achieve 1pmol specifically bound DNA-sites in the 2P-complex. After 30 min incubation at 25°C, samples were electrophoresed on 1mm 6% acrylamide-TBE-gels (acrylamide:bisacrylamide ratio 39:1, 50mM Tris borate,

pH8.3, 1.25mM EDTA) at 10V/cm. After the run, the position of the purified protein-DNA complex is visualized by exposing the gels to Kodak XAR5-films for 2h at 4°C.

Analytical mobility shift assay

Binding reactions and gel electrophoresis were carried out as described in the previous section. Reactions contained 0.4-4ng labelled sBS2-fragment, 500ng poly(dI-dC).(dI-dC) and serial protein dilutions in BB. After electrophoresis gels were dried and exposed to Kodak XAR5-films. For the quantification of TGGCA-binding activity we used dilutions of subsaturating protein amounts. Free and retarded DNA-fractions were cut out from dried gels (using the Xray-film as template) and the radioactivity was counted. Values were corrected for average recovery.

DNaseI-protection and methylation-interference experiments

Footprinting-experiments (19) were performed as follows. Standard binding reactions of sBS2-fragment (labelled at the EcoRI-site) and DCF-proteins were incubated for 30 min at 25°C. $MgCl_2$ (5mM final conc.) and DNaseI (0.2µg/assay) were added and the incubation was continued for 3 min. Samples were electrophoresed and the DNA-fractions (free and retarded DNA) were recovered by electroelution and analysed on 20% acrylamide-8M urea-gels. Sequencing reactions were carried out as described (20). For methylation-interference experiments (21), sBS2-DNA (labelled at the EcoRI-site) was first partially methylated with dimethyl-sulphate (0.2%, 20 min on ice) before it was used in mobility shift assays. Again DNA-fractions (free and retarded DNA) were recovered from native gels and after piperidine-treatment analysed on 20% sequencing gels.

SDS-PAGE analysis of PMSE-purified TGGCA protein

SDS-polyacrylamide electrophoresis was performed as described by Laemmli (22); proteins were precipitated according to Wessel and Flügge (23). Silver-staining followed the method of Merril et al (24). A gel slice containing the protein-DNA complex was cut out from the PMSE-gel, put on top of a 1mm SDS-polyacrylamide gel (18%) and overlayed with SDS-sample buffer. Reference samples were loaded when 50% of the dye front had entered the stacking gel. After electrophoresis the gel was silver-stained, dried and autoradiographed.

SDS-PAGE analysis of proteins and V8-peptides

For peptide mapping of DCF-proteins, S.aureus V8-proteolysis was performed during SDS-PAGE. Proteins were resolved on a SDS-gel and coomassie-stained. The gel was soaked in H_2O for 1h to remove the fixation solution. Stained protein bands were cut out, put into the gel slots of a second SDS-gel and overlayed with SDS-sample buffer containing 500ng V8-protease. When the dye front had reached the separation gel, the run was stopped for 1h to allow for proteolysis. After electrophoresis the gel was silver-stained.

RESULTS

Prepurification of TGGCA protein from chicken liver

The ubiquituous presence of the TGGCA protein in chicken tissues enabled us to look for the optimal starting material for its purification. We chose liver tissue for several reasons: (1) It is available in large amounts from chicken slaughter houses, (2) it can be easily dispersed resulting in good yield of nuclei, and (3) its nuclear extract shows the highest specific binding activity when compared to other chicken tissues or cell lines (data not shown).

A procedure for small scale preparation of the nuclear extract was published previously (2). To increase the specific binding activity for large scale preparations, nuclei were prewashed and differentially eluted with low-salt. A 100 mM-NaCl preeluate (PE) is almost free of specific binding activity (see Table 1). The 100mM to 300mM nuclear extract (NE) is passed over DEAE-Sephacel at 250 mM NaCl with the TGGCA binding activity being found in the flow through (DEAEFT). Subsequent purification steps included chromatography on phosphocellulose (PC), Heparin-Sepharose (HS) and calf thymus DNA-cellulose (DC), in each case eluting the TGGCA protein by linear NaCl-gradients (for details see "materials and methods"). The TGGCA-binding activity was monitored and quantified by mobility shift analysis (17, 18). In Table 1, steps 1-5, data are summarized of this prepurification scheme naming the pooled active gradient fractions according to the respective chromatographic step (e.g. PCF = phosphocellulose-fraction). The four column steps

TABLE 1 PURIFICATION OF TGGCA PROTEIN FROM CHICKEN LIVER

STEP	FRACTION	VOLUME ml	TOTAL PROTEIN mg	SPEC. ACTIVITY[b] pmol/mg	PURIFICATION -fold	PURITY %	YIELD %
-	PE	550.0	209.90	0.3	-	-	-
1	NE[a]	310.0	543.00	15.7	1.00	0.1	100.0
2	DEAEFT	550.0	407.00	15.1	0.96	0.1	72.3
3	PCF	90.0	54.00	80.8	5.15	0.57	51.2
4	HSF	22.0	9.50	428.0	27.26	3.0	47.5
5	DCF	3.5	0.67	5105.3	325.20	35.7	39.8
6	PMSEF[c]	-	0.03	14286.3	910.00	100.0	3.3

[a] 500g tissue, $2x10^{11}$ nuclei; [b] specifically bound DNA-site/total protein;
[c] TGGCA protein was calculated from complexed DNA-amount, based on dimer-binding (monomer 35kd). This step was done with part of DCF only. Presented values are adjusted to allow comparison with steps 1-5.

led to a 325-fold increase of specific activity with a yield of roughly 40% of initial binding activity.

To ensure that we monitor only TGGCA-binding activity during the purification, we generally used sBS2, a 100 bp fragment containing two consensus recognition sequences for the TGGCA protein (6; its sequence is depicted in Figure 3) for mobility shift analysis. However, we want to stress that we did not observe any difference when we used either a monomer of the synthetic binding site or DNA-fragments containing natural TGGCA-binding sites from the 5'-flanking region of the chicken lysozyme gene (2).

The electrophoretic mobilities of specific protein-DNA complexes formed between the TGGCA protein and sBS2-fragment after each purification step are shown in Figure 1. The retarded bands 1P and 2P represent one and two TGGCA protein complexes bound to the sBS2-fragment. Interestingly, nuclear extract (lanes 2-4) and DEAEFT (lanes 6&7) show a third complex above the 2P-position. We have strong evidence that this is neither due to another DNA-binding protein nor to unspecific binding of the TGGCA protein itself. It is most likely caused by a protein interacting with the TGGCA protein via protein:protein interaction. Only a very small part of this activity, which by itself does not bind to the sBS2-fragment, copurifies with the TGGCA protein (R. Rupp, unpublished data).

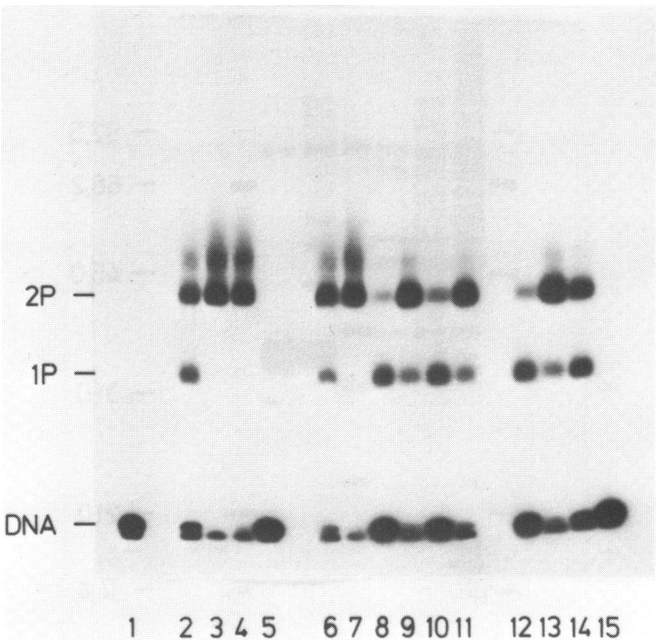

Fig. 1. Mobility shift analysis of the TGGCA protein during purification.
Labelled sBS2-fragment (0.4ng,10000cpm) was incubated with two appropiate protein dilutions from each step of the purification to yield similar saturation-levels of the binding reactions. Samples were analysed on a 6% acrylamide-gel. Lane 1: DNA, no protein. The DNA was incubated with: NE (lanes 2-5), DEAEFT (lanes 6&7), PCF (lanes 8&9), HSF (lanes 10&11) and DCF (lanes 12&15). DNA-competition assays were performed with the higher protein concentrations of NE and DCF (as in lanes 3&13 respectively) in presence of 50ng of either EcoRI-linearized pBR322 (lanes 4&14) or unlabelled sBS2-fragment (lanes 5&15). The position of the two major retarded bands represent protein-DNA-complexes of one (1P) or two TGGCA protein complexes (2P) bound to the sBS2-fragment.

DNA competition experiments with the nuclear extract (Figure 1, lanes 4&5) and the DC-fraction (lanes 14&15) clearly demonstrate the conservation of TGGCA-binding specificity. The slight competition observed with pBR322 DNA (lane 14) is due to a number of sites on the plasmid with low homology to the TGGCA motif. More important, Figure 1 shows that the mobilities of the 1P- and 2P-complexes are identical throughout the purification (including the small amount of TGGCA-binding activity in

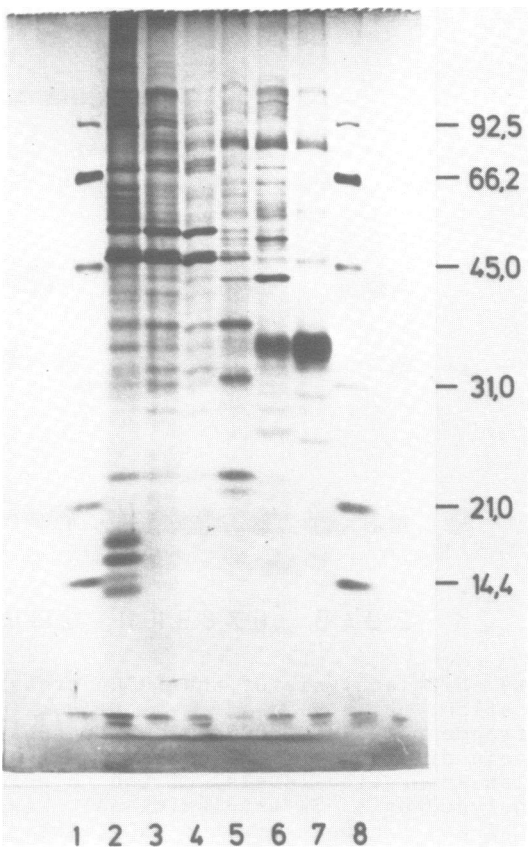

Fig. 2. SDS-polyacrylamide gel electrophoresis of the purification fractions.
Protein (10µg) from each step of the purification was electrophoresed on a 15% SDS-acrylamide gel. Lanes 1&8: marker proteins (numbers to the right giving the molecular weights in kd). Lane 2: nuclear preeluate, PE. Lane 3: 100-300mM nuclear extract, NE. Lane 4: DEAE-Sephacel flow-through, DEAEFT. Lane 5: phosphocellulose-fraction, PCF. Lane 6: Heparin-Sepharose-fraction, HSF. Lane 7: DNA-Cellulose-fraction, DCF. After electrophoresis the proteins were visualized by silver-staining.

the pre-eluate; data not shown). This argues strongly against proteolysis during purification.

Figure 2 shows the protein composition of the various purification fractions on SDS-PAGE. Although a protein, appearing as a broad band of approximately 35 kd, was prominently en-

riched in the DC-fraction, it still contained a large variety of minor proteins of different molecular weight. Activity calculations based on the binding of a TGGCA protein dimer per DNA-binding site (6, 8, 25, 26) and a molecular weight of 35 kd for the protein monomer result in 35.7% purity of this fraction. This value is well in agreement with SDS-PAGE analysis (Figures 2 and 5A).

Specific interaction of the TGGCA protein with a synthetic DNA-binding site

The DNA-sequence requirements of binding sites for NF-I have been studied in great detail (26, 27). All these results confirm our early conclusion (2, 6) that the optimal recognition sequence is a palindromic motif of 5'-YTGGCA-3' separated by 3 freely variable base pairs as we use it here in our synthetic binding site.

To study the interaction of the TGGCA protein with the sBS2-fragment (Figure 3), we combined the mobility shift assay with DNaseI-footprinting (19) or methylation interference studies (21). Asymmetrically labelled sBS2-fragment was incubated with increasing amounts of DC-fraction and partially digested with DNaseI. The protein-DNA complexes were separated from free DNA on a native acrylamide gel. We isolated both free and retarded DNA and analysed it parallel to chemical sequencing-reactions (20) on a denaturing gel (Figure 3A). Depending on the protein amount (lanes 7-10), we observed two footprints over the binding sites in the retarded DNA, each extending approximately 25 bp.

For the analysis of contactpoints made by TGGCA protein we incubated partially methylated sBS2-fragment with the DC-fraction. Free and bound DNA was separated on a native gel, as seen in Figure 1. Both DNAs were eluted and after G-specific chemical cleavage analysed on a denaturing gel (Figure 3B). Clearly, methylation of the G-residues at positions -5/-4 and +3 in respect to the axis of symmetry of the palindrom interferes with binding and their respective cleavage products are depleted from the bound and enriched in the nonbound DNA-fraction (Figure 3B, lanes 14&15). In contrast, methylation of guanines

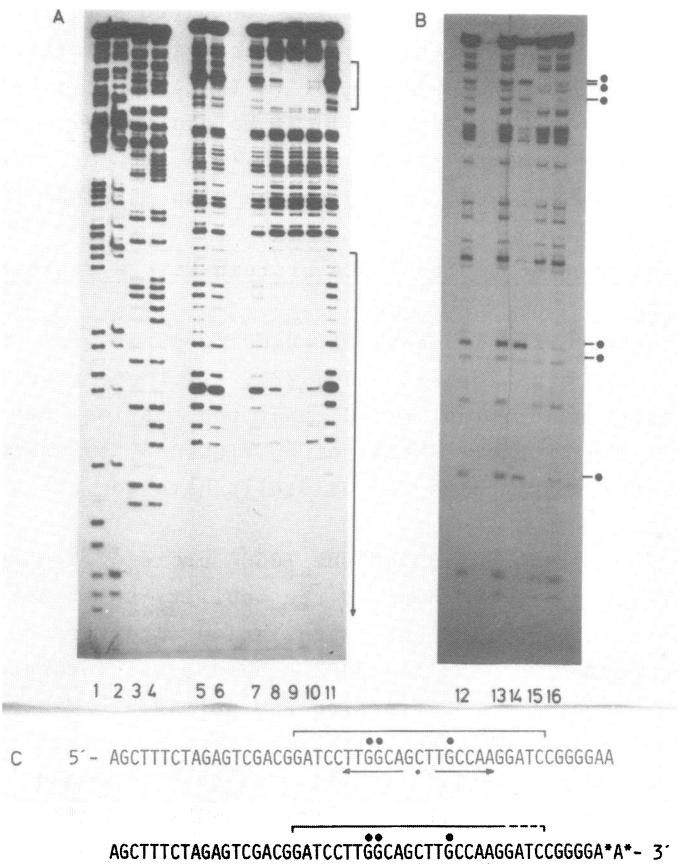

Fig. 3. Specific interaction of TGGCA protein and sBS2-DNA. The sBS2-fragment (labelled at the 3'-end of the EcoRI-site; 4ng, 85000cpm) was incubated with increasing amounts of DCF (0.03-0.25µl). The protein-DNA interaction, giving rise to 1P/2P-complexes (Figure 1), were analysed by:
A. DNaseI-footprinting. After incubation, samples were digested with DNaseI and loaded on a native 6% acrylamide gel. The free DNA and the DNA from protein-DNA complexes 1P and 2P was eluted and analysed on a 20% sequencing gel. Lanes 1-4: A-, G-, C-, T-specific chemical sequencing reactions. DNaseI-digested samples: free DNA (lane 5,6 and 11), DNA from 1P-complex (lane 7), DNA from 2P-complexes (lanes 8-10). Brackets mark the DNA-regions protected by TGGCA protein.
B. Methylation-interference analysis. Partially methylated DNA was incubated with a saturating amount of DCF. Free DNA and protein-complexed DNA were again eluted from a native acrylamide gel and after piperidine cleavage analysed on a 20% sequencing gel. Lane 12: G-reaction, lane 13: unincubated DNA, lane 14: free DNA-fraction, lane 15: DNA from 2P-complex, lane

16: DNA from 1P-complex. Filled circles mark G-residues, which interfere with binding of the TGGCA protein.
C. Nucleotide sequence of sBS2-DNA. Arrows and dots below DNA sequence mark the palindromic consensus recognition sequences of the TGGCA protein; brackets and filled circles above refer to the DNaseI-protected regions and G(N7)-contact points. Asterisks at 3'-adenine residues indicate the labelled EcoRI-site.

at position -1 or outside the recognition sequence did not interfere with TGGCA protein binding.

The results of these experiments are summarized in Figure 3C. The fact that both footprints and the methylation-interference pattern are albeit weaker but already visible in the 1P-position (Figure 1, analysed in Figure 3, lanes 7 and 16) supports the interpretation that the first retarded DNA-fraction is derived from a protein DNA complex in which only 50% of each of the two DNA sites is occupied by TGGCA protein. Because of the perfect symmetry of the binding sites we present this analysis for one strand only, assuming that it will be identical for the other strand.

The methylation interference pattern is completely compatible with that one seen for NF-I on the origin of replication DNA of adenovirus 2 (26). It is therefore established that the protein-DNA interaction seen in band 2P of Figure 1 is due to two identical complexes of TGGCA protein with its unique mode of DNA-recognition.

Final purification of the TGGCA protein by preparative mobility shift electrophoresis (PMSE)

Although our purification led to a significant enrichment in TGGCA-binding activity, it did not yet allow the unambiguous identification of the corresponding protein species. For this purpose we developed a novel kind of electrophoretic purification procedure. The native protein-DNA complex is isolated by preparative mobility shift electrophoresis (PMSE-fraction; Table 1, step 6) from a protein-DNA incubation mixture containing a 250-500fold higher concentration of both components than normally used in analytical mobility shift assays.

The protein component of the electrophoretically purified complex was identified by subsequent SDS-PAGE and silver-stain-

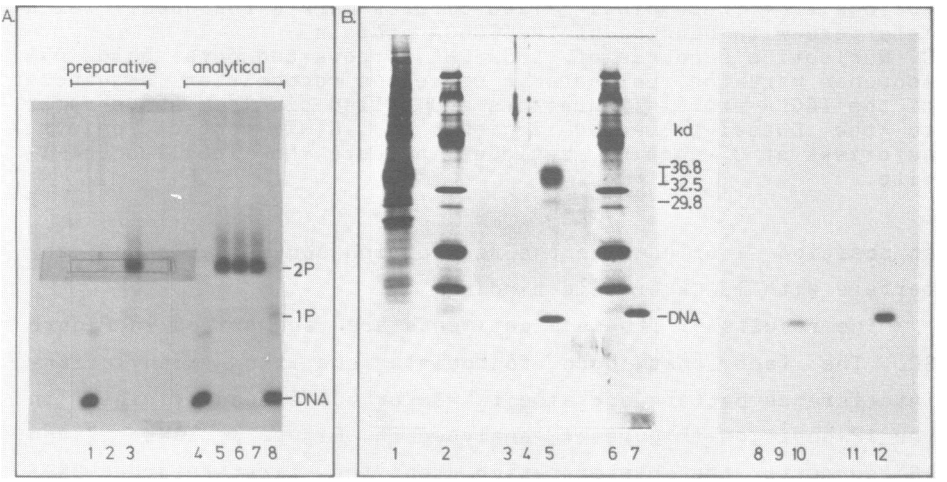

Fig. 4. Mobility shift purification of the TGGCA protein.
A. Preparative mobility shift electrophoresis (PMSE). sBS2-fragment (128ng, 4pmol TGGCA-binding sites, 15 000ccpm) was incubated with DCF (1.56µg total protein, 7.95pmol TGGCA-binding activity) and loaded on a 6% acrylamide-TBE-gel. Lane 1: DNA, lane 2: protein, lane 3: DNA & protein. Lanes 4-8 represent analytical mobility shift assays (similar to Figure 1, lanes 12-15), including DNA-competition with pBR322 (lane 7) and sBS2-fragment (lane 8).
B. SDS-PAGE of 2P-complex. A gel slice, containing the preparative 2P-complex plus the corresponding region of the control lanes (Figure 4A, lanes 1-3), was cut out and put on top of a SDS-acrylamide gel. After the run both proteins and DNA were silver-stained. Lanes 3-5 represent lanes 1-3 of Figure 4A, lanes 2&6 show marker proteins. Lane 1: DCF (1.56ug), lane 7: sBS2-fragment (128ng). Molecular weights of TGGCA proteins present in lane 5 are given in kd. Lanes 8-12 show the autoradiograph of part (lanes 3-7) of the dried SDS-acrylamide gel.

ing. For the molecular weight determination we set the PMSE-scale to 1pmol specifically bound DNA-site, which allows the identification of even small protein species ($\geq$10kd) in one slot of gels used in standard analytical assays.

Figure 4 documents an experiment in which the prepurified TGGCA protein of the DC-fraction is purified to completion via its specific complex with the sBS2 DNA-fragment. Figure 4A shows the autoradiograph of the preparative electrophoretic separation of the protein-DNA complex (lane 3). For comparison we ran next to it standard analytical assays, which demonstrate

the identity of the retarded 2P bands. The controls (run in slots 1&2) contain solely the DNA component (both double-stranded poly(dI-dC) and sBS2-fragment) or the protein component (DCF) respectively, in same amounts as combined in the complete incubation mixture for the formation of the TGGCA protein-DNA complex. For the SDS-PAGE analysis of the protein component in the retarded DNA complex, a gel slice was cut out covering the whole width of lanes 1-3 (Figure 4A) at the height of the 2P signal. It was put on top of a standard SDS-acrylamide gel. Figure 4B shows the silver-stain of the protein-gel, with lanes 3-5 representing the transferred material of slots 1-3 in Figure 4A. Together with size-markers (lanes 2&6) the SDS-gel was run with the same controls as the preparative mobility shift electrophoresis, i.e. the same amounts of DCF (Figure 4: A,lane 2 and B, lane 1) and of sBS2-DNA plus double-stranded poly(dI-dC) (Figure 4: A, lane 1 and B, lane 7).

The absence of any signal in lanes 3 and 4 of Figure 4B proves that no DNA and no protein is running by its own to the position of the 2P-complex in the native gel. Therefore the bands present in lane 5, Figure 4B must represent the two components of the 2P-complex, i.e. the sBS2-fragment and the TGGCA protein. We checked previously that SDS disrupts the protein-DNA complex (data not shown). Autoradiography of the dried SDS-protein gel (Figure 4B; lanes 8-12 cover lanes 3-7 of the silver-stain) identifies the naked DNA-fragment as the fastest migrating component. Moreover, the densitometric scanning of both signals can be used to quantify the efficiency of the gel slice transfer, which will be discussed later.

To our surprise we detected more than one protein species in the specific TGGCA protein-DNA complex (Figure 4B, lane 5). There were several proteins, whose range in molecular weight (indicated between lanes 7 and 8) covers the DCF-protein being most enriched during purification (Figure 1). Since we have carefully characterized the protein-DNA interaction using the retarded protein-DNA complex (results of Figure 3), all these proteins must bind the DNA by the characteristic mode of TGGCA protein/NF-I.

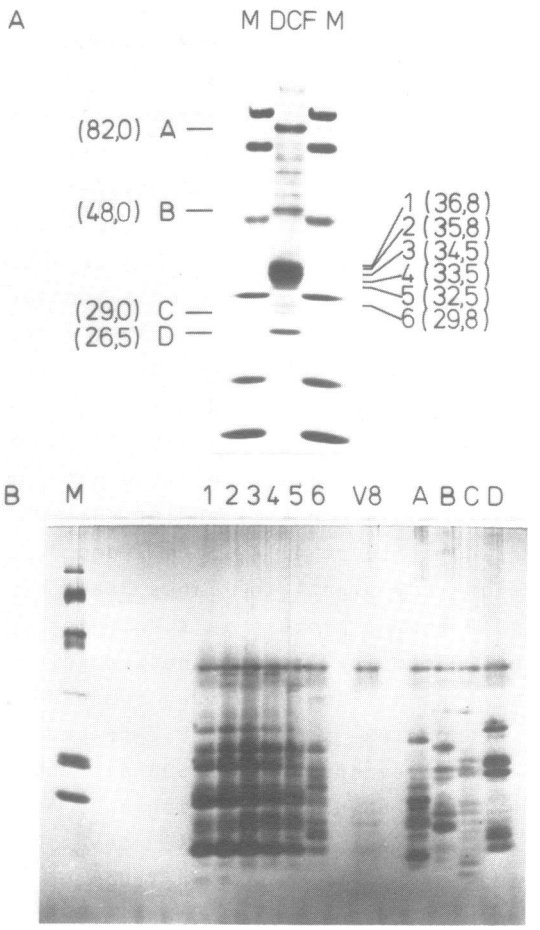

Fig. 5. Peptide mapping of TGGCA proteins by limited proteolysis and SDS-PAGE

A. DCF-proteins (35µg) were electrophoresed on a 1mm 18% SDS-acrylamide gel and coomassie-stained. Molecular weights of protein contaminants (A-D) and TGGCA proteins (1-6) are given in kd. M: marker proteins.

B. Silver-stain of DCF-peptides. Single protein bands (Figure 5A) were cut out and put into the gel slots of a 20% SDS-acrylamide gel. The gel pieces were overlayed with 10µl sample buffer containing 500ng S.aureus V8-protease. M: marker proteins, V8: 500ng protease; numbers 1-6 and letters A-D refer to proteins as indicated in part A.

Peptide mapping of TGGCA proteins

On high percentage SDS-gels, we were able to identify 6 TGGCA protein species (Figure 5A, numbered 1-6). Five of them form

the prominent set of DCF-proteins (Figure 1). They range in size from 36.8 to 32.5kd, showing a somewhat regular 1kd difference between each of them. The sixth protein species has an apparent molecular weight of 29.8kd and is represented by the upper band of a weaker doublet (Figure 4B, lanes 1&5 and Figure 5A). The proteins are not present in equimolar amount, the species of 35.8kd and 34.5kd being most abundant. The protein composition of the DC-fraction, including the relative abundance of the 6 TGGCA proteins, proved to be stable in several preparations (data not shown).

To get insight into the relationship of the TGGCA protein species we compared their peptide pattern after partial V8-digest during SDS-PAGE. As a control we included the peptide analysis of the prominent protein-contaminants present in the DC-fraction (Figure 5B). The slowest migrating protein component present in all slots was the V8-protease itself (see reference-slot). Clearly all TGGCA protein species share most of their proteolytic peptides. While some of the small differences might be explained by a different completion of the partial V8-digest, some might be due to differences in the protein itself. All the copurified protein-contaminants in DCF show a significantly different peptide pattern. From this we conclude that the specifically DNA-binding TGGCA protein comprises a family of proteins with closely related primary structure.

DISCUSSION

Purification of the TGGCA protein

We have optimized the method of nuclear extract preparation for the TGGCA protein. Differential low-salt elution of prewashed nuclei results in 110mg of nuclear protein from 100g fresh chicken liver tissue (Table 1), which corresponds to a roughly 165-fold prepurification of the TGGCA protein with respect to total cellular protein (p.92 in ref. 28). The following purification by 4 chromatography steps increases specific TGGCA-binding activity 325-fold (Table 1), as quantitated from analytical mobility shift assays. This is reflected by enrichment of a set of proteins with an average molecular wheight of 35kd on

SDS-gels (Figure 2). On the basis of two 35kd protein monomers binding to one palindromic DNA binding site, we calculate a value of 36% purity from the specific activity of the DC-fraction. The final purification to a homogeneous protein fraction was achieved by PMSE. In toto, this means an approximate 150.000-fold purification of the TGGCA protein. We estimate from the data of Table 1 that roughly 50.000 TGGCA protein monomers can be eluted per nucleus of chicken liver cells.

PMSE - a simple electrophoretic procedure to characterize the protein moiety of specific DNA-binding proteins.

Preparative use of the mobility shift assay resulted in the final purification of the TGGCA protein. This was facilitated on one hand by the surprising finding that no unbound proteins were detected at the position of the TGGCA protein-DNA complex. Running protein samples without DNA on mobility shift gels showed that the proteins do not enter the gels at all or stay close to the slot edge (data not shown). On the other hand, careful titration of the two components (protein and DNA) was of the same importance, because it allowed us to increase drastically their concentrations without losing either specific complex formation or its clear resolution on native polyacrylamide gels. In combination with SDS-PAGE, we employed PMSE to identify the molecular weight of the TGGCA protein. We used a 2-fold excess of TGGCA-binding activity over DNA-binding sites (details see "materials and methods" and legend to Figure 4). After densitometric scanning of the autoradiographs (Figure 4A, lane 3 and Figure 4B, lane 5) we found 25% of the input DNA in the 2P-complex with no apparent loss during transfer of the gel slice and SDS-PAGE. The calculated yield of TGGCA protein (12.5%) present in the 2P-complex corresponds well with the amounts of stained proteins seen in the total and in the purified protein fraction respectively (Figure 4B, lanes 1 and 5).

Besides the fact that PMSE is based on widely used techniques only, it offers several advantages. Site-specific DNA-affinity chromatography still requires identification of the protein species corresponding to the specific DNA-binding activity (29-31). PMSE makes this step unnecessary, because the proteins are purified by the same procedure as used for their

characterization. This method will be equally applicable to the purification of other DNA-binding proteins after partial pre-purification (≥1% purity). Other advantages of this procedure are that it very specifically discriminates in one step between different protein-DNA complexes and that it can be used for the isolation of protein components binding to protein-DNA complexes only via protein:protein interaction.

Microheterogeneity of the TGGCA protein

Our analysis of the protein-DNA interaction in the mobility shift assay (Figure 3) unambiguously identifies the TGGCA protein by both DNaseI footprinting and methylation interference (2, 6, 8, 26; U. Borgmeyer, unpublished results). The high evolutionary conservation of the mechanism of DNA-recognition between HeLa cell NF-I and the chicken TGGCA protein is reflected by the results of the protein-DNA contactpoint analysis. N7 methylation of the first G-residue of the 5´-TGG-3´ triplet interferes stronger with protein binding than N7 methylation of the second G-residue (Figure 3, lanes 14+15). The same intimate characteristic was observed for NF-I (Figure 3 in ref. 26; functional significance in Figure 5 of ref. 27). Both proteins share a second activity, i.e. enhancement of the initiation reaction of Adenovirus replication **in vitro** (published for HSF in ref. 4). Since our DC-fraction is also active in this respect (data not shown), we are purifying a biologically active protein.

The chicken liver TGGCA protein purified to homogeneity by preparative mobility shift electrophoresis (PMSE) yields a family of 6 protein species. Their slight variations in apparent molecular weight (36.8kd to 29.8kd), and possibly the formation of mixed protein dimers prevent the detection of different protein-DNA complexes in native gels (Fig.1 and Fig.4). From DNA competition and methylation interference experiments, we conclude that all species interact identically with the TGGCA-binding site. Otherwise we would not observe an almost complete depletion of distinct methylated G-residues from the 2P-complex in methylation interference experiments (Figure 3B, lane 15). Additionally, partial V8-protease digests revealed a nearly

identical peptide pattern for the various TGGCA protein spe-
purification of a family of NF-I proteins with molecular weights from 66kd to 52kd. In contrast to the TGGCA protein family this larger mass difference causes multiple protein-DNA complexes in mobility shift assays (Fig.2c in ref. 16). Although we cannot rigorously exclude proteolysis during purification to be responsible for both the lower molecular weight of the TGGCA proteins and their microheterogeneity, several lines of evidence argue against this: (1) The analytical mobility shift analysis shows an identical pattern for all purification steps. (2) All tested nuclear extracts from chicken tissues or cell lines showed the same mobility shift pattern (R. Rupp, U. Borgmeyer, T. Grussenmeyer, unpublished results) and (3) the purification conserves besides specific DNA-binding the activity to stimulate Adenovirus replication **in vitro**. Though it is possible that the microheterogeneity of TGGCA protein is the result of physiological proteolysis in the living cell, it seems to be more likely that the various protein species are either the result of different posttranslational modifications or are coded by slightly differing mRNAs. A similar microheterogeneity was recently published for purified AP-I, another transacting factor (32), and was reported by M.Krasnow for the products of the Drosophila Ubx gene (cited in ref. 33).

We found two TGGCA-binding sites 6.1kb upstream of the transcriptional start site of the chicken lysozyme gene (1, 2) within a chromatin DNase-hypersensitive site strictly correlated to gene activity (10). This hypersensitive DNA region harbors a cellular enhancer element (11), the full cell-specific activity of which is dependent on funtional binding sites for the TGGCA protein (Theisen et al., in preparation).

An increasing number of publications demonstrate either posttranslational modifications of transacting factors (30, 32, 34-38) or selective binding of different DNA-binding proteins cies, reflecting a high similarity in primary structure.

Several groups purified the human counterpart NF-I from HeLa cells and published different molecular weights of the protein. The data range from 160kd (25) to 47kd (3) for a single protein species. Interestingly, Rosenfeld and Kelly (15) reported the

to identical or overlapping target sequences (39, 40) to be involved in differential gene activation. It is possible that selective action of individual TGGCA protein species might be responsible for its various functional specificities. This possibility is even more likely in view of the observation that 2D-electrophoresis (NEPHGE/SDS-PAGE) splits TGGCA proteins in two populations of different isoelectric points (R.Rupp, unpublished).

ACKNOWLEDGEMENT

We would like to thank Dr. Joachim Nowock for help in the initial phase of the project, Uwe Borgmeyer for stimulating discussions and Gaik Pin Wee for secretarial assistance. The work was funded by grants from the Bundesministerium für Forschung und Technologie (BCT 0381/5) and Fonds der Chemischen Industrie.

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. Nagata, K., Guggenheimer, R.A., Enamoto, T., Lichy, J.H. and Hurwitz, J. (1982) Proc.Nat.Acad.Sci.USA, 79, 6438-6442.
4. Leegwater, P.A.J., van der Vliet, P.C., Rupp, R.A.W., Nowock, J. and Sippel, A.E. (1986) EMBO J., 5, 381-386.
5. Rawlins, D.R., Rosenfeld, P.J., Wides, R.J., Challberg, M.D. and Kelly, T.J. (1984) Cell, 37, 309-319.
6. Nowock, J., Borgmeyer, U., Püschel, A.W., Rupp, R.A.W. and Sippel, A.E. (1985) Nucleic Acids Res., 13, 2045-2061.
7. Schneider, R., Gander, I., Müller, U., Mertz, R. and Winnacker, E.L. (1986) Nucleic Acids Res., 14, 1303-1317.
8. Nagata, K., Guggenheimer, R.A. and Hurwitz,J. (1983) Proc.Nat.Acad.Sci.USA, 80, 6177-6181.
9. Hay, R.T. (1985) J.Mol.Biol., 186, 129-136.
10. Sippel, A.E., Fritton, H.P., Theisen, M., Borgmeyer, U., Strech-Jurk, U. and Igo-Kemenes, T. (1986) in Grodzicker, T., Sharp, P.A. and Botchan, M.R. (eds.), DNA Tumor Viruses: Gene Expression and Replication. Cancer Cells 4/1986, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
11. Theisen, M., Stief, A. and Sippel, A.E. (1986) EMBO J., 5, 719-724.
12. Buetti, E. and Kühnel, B. (1985) J.Mol.Biol., 190, 379-389.
13. Miksicek, R., Borgmeyer, U. and Nowock, J. (1987) EMBO J., 6, 1355-1360.

14. Cordingley, M.G., Riegel, A.T. and Hager, G.L. (1987) Cell, 48, 261-270.
15. Rosenfeld, P.J. and Kelly, T.J. (1986) J. Biol. Chem., 261, 1398-1408.
16. Jones, K.A., Kadonaga, J.T., Rosenfeld, P.J., Kelly, T.J. and Tjian, R. (1987) Cell, 48, 79-89.
17. Fried, M. and Crothers, D.M. (1981) Nucleic Acids Res., 9, 6505-6525.
18. Garner, M.M. and Revzin, A. (1981) Nucleic Acids Res., 9, 3047-3060.
19. Galas, D.J. and Schmitz, A. (1978) Nucleic Acids Res., 5, 3157-3170.
20. Maxam, A. and Gilbert, W. (1980) Meth. in Enzymol., 65, 499-560.
21. Siebenlist, U., Simpson, R.B. and Gilbert, W. (1980) Cell, 20, 269-282.
22. Laemmli, U.K. (1970) Nature, 227, 680-685.
23. Wessel, D. and Flügge, U.I. (1984) Anal. Biochem., 138, 141-143.
24. Merril, C., Goldman, G. and van Keuren, M. (1982) Electrophoresis, 3, 17-23.
25. Diffley, J.F.X. and Stillman, B. (1986) Mol. Cell. Biol., 6, 1363-1373.
26. de Vries, E., van Driel, W., van den Heuvel, S.J.L. and van der Vliet, P.C. (1987) EMBO J., 6, 161-168.
27. Rosenfeld, P.J., O'Neill, E.A., Wides, R.J. and Kelly, T.J. (1987) Mol. Cell. Biol., 7, 875-886.
28. Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J.D. (1983) Molecular Biology of the Cell. Garland Publishing, New York.
29. Briggs, M.R., Kadonaga, J.T., Bel, S.P. and Tjian, R. (1986) Science, 234, 47-52.
30. Lee, W., Mitchell, P. and Tjian, R. (1987) Cell, 49, 741-752.
31. Montminy, M.R. and Bilezikjian, L.M. (1987) Nature, 328, 175-178.
32. Engel,P., Imagawa, M., Chin, R., Stein, B., Imbra, R.J., Rahmsdorf, H.J., Jonat, C., Herrlich, P. and Karin, M. (1987) Cell, 49, 729-739.
33. Roberts, M. (1987) Nature, 327, 556-557.
34. Sen, R. and Baltimore, D. (1986) Cell, 47, 921-928.
35. Atchinson, M.L. and Perry, P.R. (1987) Cell, 48, 121-128.
36. Hayes, T.E., Kitchen, A.M. and Cochran, B.H. (1987) Proc. Nat.Acad.Sci.USA, 84, 1272-1276.
37. Kingston, R.E., Schuetz, T.J. and Larin, Z. (1987) Mol. Cell. Biol., 7, 1530-1534.
38. Zimarino, V. and Wu, C. (1987) Nature, 327, 727-730.
39. Staudt, L.M., Singh, H., Sen, R., Wirth, T., Sharp, P.A. and Baltimore, D. (1986) Nature, 323, 640-643.
40. Barberis, A., Superti-Furga, G. and Busslinger, M. (1987) Cell, 50, 347-359.